An Accurate and Efficient Procedure for Estimating Process Design Reliability

Richard J. MacDonald

BSChE: University of Kansas 1965
MSChE: University of Kansas 1986

Submitted to
the Department of Chemical and Petroleum Engineering
and the Faculty of the Graduate School of the University of Kansas
in Partial Fulfillment of the Requirements for degree of
Doctor of Philosophy

Dissertation Committee

Colin S. Howat III, Chairman

George W. Swift

Kenneth A. Bishop

Brian E. Thompson

William G. Bulgren

November, 1992

Abstract

Safety factors are an expensive, but necessary part of any chemical engineering process design. Unfortunately, safety factor selection procedures are often little more than approximate rules of thumb based on experience. Increasing economic pressures and the risk of failure demand a rigorous procedure to size safety factors for maximum benefit at minimum cost. This requires statistical descriptions of the uncertainties and a method for calculating design reliability (odds of success).

Hybrid distributions (possibility and/or probability distributions) are the best (albeit undeveloped) tools for describing uncertainty. However, calculation of reliability is prohibitively expensive, requiring thousands of process simulations for adequate results. This research presents a new calculation method that is orders of magnitude better than previous methods. It provides accurate and precise reliability estimates for just tens or hundreds of process simulations. This includes essentially free and unlimited sensitivity analyses.

The key is to mathematically approximate the constraint boundary (border separating the regions of success and failure) independent of the statistics. This is done by finding and geometrically interpolating between points on the boundary. The statistical calculations can then be rapidly performed using the mathematical boundary-approximation. Process simulation effort is spent finding the boundary points instead.

While applicable to hybrid distributions the procedure was developed and tested with probability distributions only. The test problem was binary and ternary distillation with feed flowrate, tray efficiency, and thermodynamic database uncertainties. The procedure was shown to be orders of magnitude faster than Conventional Monte Carlo.

Computational requirements increase with the number of uncertain parameters, but are reasonable and remain much lower than for Conventional Monte Carlo. The reliability estimate may be slightly inaccurate—usually within 0.5% and always within 1%-2%—however, this error can be estimated and compensated for with some additional effort.

Distillation control variables were eliminated by implicit substitution. Other process operations may not allow this simplification.

Future work should extend the procedure to hybrid statistics, provide for explicit handling of control variables when necessary, and optimize the procedure. Most importantly, better statistical descriptions of chemical engineering process design uncertainties are required.

Summary

A process designer is responsible for specifying a piece of equipment to perform some given task, such as a chemical separation or reaction, and guaranteeing that it will work. This can be a very difficult task because of the inherent uncertainties. For example, the designer may have to account for:

1) Uncertainties in the future operating conditions and economic factors

2) Process model and database inaccuracies and uncertainties

The designer must compensate by adding safety factors. Since this is expensive, safety factor selection should be optimized to provide maximum benefit for minimum cost. Unfortunately, there is often no rigorous procedure available, so the designer must rely on experience and rules of thumb. However, this is potentially wasteful and does not guarantee success.

A formal analysis of the risks and uncertainties would allow optimal sizing of the safety factors. This requires statistical tools to describe the uncertainties and a method to calculate the design reliability (odds of success). However, current capabilities are severely limited.

One option is to use minimum/maximum bounds for each uncertain parameter. However, this provides very weak information and there is no estimate of design reliability—either the design *will* work, or it *might* fail. Also, additional information cannot be used. Another option is to use probability distributions to describe the uncertainties, using numerical integration or Conventional Monte Carlo to calculate the design reliability. However, the former method is limited to simple problems and the latter is computationally expensive, requiring hundreds of process simulations for adequate precision. Also, probability distributions require substantial amounts of information to be accurately determined, and thus are inadequate for highly uncertain parameters, e.g., future operating conditions and systematic errors in the process models.

The best statistical tools are “hybrid” distributions, which use probability and/or possibility (“fuzzy”) distributions. Possibility distributions are complementary to probability distributions, being able to handle minimum/maximum bounds as well as qualitative information such as “about” 50%. However, possibility statistics is still a young theory that is undeveloped in chemical engineering. Also, the design reliability calculation procedure for hybrid statistics requires thousands of process simulations for adequate results.

A final difficulty is that statistical distributions are only mathematical tools for describing uncertainties, resulting in unavoidable assumptions and uncertainties-in-the-uncertainties. A sensitivity analysis can determine if these assumptions are significant by altering the

distributions and recalculating the design reliability. However, if each calculation is lengthy, the complete analysis can be prohibitive.

This research presents a new procedure that breaks these computational barriers. Accurate and precise design reliability estimates can be obtained for just tens or hundreds of process simulations. This includes essentially free and unlimited sensitivity analyses.

The procedure mathematically approximates the constraint boundary (border separating the region of success and failure) independent of the statistics. This is accomplished by finding a number of points on the constraint boundary, then geometrically interpolating between them. The statistical calculations (reliability estimation and sensitivity analyses) can then be performed very rapidly using the mathematical boundary-approximation instead. Each boundary point requires several process simulations and therefore incurs a computational cost. However, by starting with a rough initial boundary-approximation, then obtaining boundary points only in the most statistically important locations, the boundary-approximation can be converged with a minimum of computational effort.

For this research, the boundary-approximation procedure was developed and tested with probability statistics only. Extension to hybrid statistics is straightforward, however. The procedure was demonstrated with binary and ternary distillation process operations. Uncertainties were assumed in the feed flowrates, tray efficiency, and thermodynamic database.

The new procedure is orders of magnitude superior to Conventional Monte Carlo. For the same accuracy, it was 20–7000 times faster than the 95% confidence interval of Conventional Monte Carlo. Alternatively, for the same computational effort it was 10–100 times more accurate than the 95% confidence interval of Conventional Monte Carlo, and is 4–30 times more likely to be accurate than a Conventional Monte Carlo calculation.

The boundary-approximation procedure requires a small initialization effort equivalent to 10-60 Conventional Monte Carlo simulations (increasing with the number of uncertain parameters). Initial accuracy is usually within 1.0% and at worst 5.0%. The procedure then converges much faster than Conventional Monte Carlo to a final accuracy usually within 0.0%–0.5%, with a worst case of 1.5%.

This potential for inaccuracy is a minor limitation of the boundary-approximation procedure that can be reduced by further work. If very high accuracy is critical, error can be estimated and compensated for by combination with Conventional Monte Carlo.

As the number of uncertain parameters (dimension) increases, the boundary-approximation procedure requires more initialization effort and may exhibit slower convergence. Thus its superiority over Conventional Monte Carlo may diminish for large dimensions. However, this was not encountered during testing up to 10 dimensions. Also,

slow convergence is relatively unimportant because of the excellent initial accuracy (1.0% – 5.0%). Finally, dimension can be minimized by parameter screening to identify and eliminate the statistically insignificant parameters.

The presence of process control variables significantly complicates the procedure. Fortunately, distillation control variables can be eliminated by substitution, so the boundary-approximation procedure was not developed to handle them explicitly. However, other process operations may not allow this simplification, so the procedure should be extended to handle this.

The procedure can be applied to multi-operation flowsheets by treating the flowsheet as a single process operation. However, this may not be the most efficient approach because of the large dimensionality. Alternative treatments should be explored.

Work should continue along two parallel paths:

1) The boundary-approximation procedure should be improved in the following areas:
   a) Hybrid statistics
   b) Testing with higher dimensions
   c) Explicit handling of control variables
   d) Efficient multi-operation flowsheet analysis
   e) Optimization of several factors to improve computational efficiency and accuracy

2) A design reliability estimate is only as good as the statistical information on which it is based. The critical path is now to develop better statistical descriptions of the uncertainties inherent in chemical engineering process design.

Table of Contents

Appendices

Acknowledgements

I am prepared to stake 95% of my reputation that 90% of this work is correct. If you want anything more definite, we need more data, more money and more doughnuts and coffee breaks.

Thanks to Chip Howat, the “smartest guy in Kansas”: Sorry I took so long to get done. I know what it cost you (no grey hairs at least). But seeing as how I worked on 2 research topics, 2 Comprehensives, 3 lawsuits, 2 process designs (1 not working), ultimate frisbee, bar band critic…

Thanks to Julie Howat: I can’t wait till you’re giving *him* extra seconds. Maybe we could do a 25 mph average next time? Start training.

Thanks to Bill Swift, my first mentor: Enjoy your retirement and expose those ethanol shysters for what they are—a net loss (thermodynamically speaking). A suggestion for a new project: Converting ethanol to Stoli.

Thanks to Ken Bishop for all your support over the years. The jury is still out as to whether computers are blessing or torture. It sure is tough out on the leading edge though. Maybe we’ll finally just run out of all possible things that can go wrong.

Thanks to Brian Thompson and Bill Bulgren for reading this thing. If its any consolation, it couldn’t possibly have been as painful as writing it.

Thanks to William Kubic for his helpfulness and encouragement during the difficult time of the first Comprehensive. By confirming a crazy idea, he gave me courage when I was tetering on the edge, and convinced me I was on the right track. Now if my execution had just been better…

Thanks to my tolerant Dad who paid a lot of the bills and made my computers possible: May your putts never break and your hands never yip (are yips hereditary or learned through observation?).

Thanks to my Mom for her nursing when the bottom fell out during that second Comprehensive. Coffee, ice cream, and pound cake may be staples, but you’re the chicken soup of my life.

Thanks to Helen, my biggest fan and biggest critic (watch what you say about me if you know what’s good for you—*you* can’t say that; only *she* can say that): Maybe I’ll finally get a chance to wear all those smart clothes gathering dust in the closet. As for my polo shirt collection of every color in the rainbow and then some: Please no more turquoise; I’ve got two already.

Now do I wear the white shirt with the blue shorts or vice-versa? If only you knew how those outfit combinations *really* got worn. P.S. About the multi-colored red (ugh!) beard—Tough, its hereditary.

Thanks to all my roomates over the years—fraternity guys, Aheem, Tom, Little Jimmie, Pat, Molly, Waldo, Kris, Cindy, Kevy, Joan, Andy, Grant, Brad, Jen, Annie, Steve, Jonathan, and Sean. I lived at 10 places over 10 years (3 in one year!). Boo-hiss to the landlords. Apologies to my long-suffering girlfriends.

Thanks to the cats—Garrison, Axel, Moose, Freaker, Stinkie, Willie, Siler, Ely and Tuscon. Special thanks to those people in dogs' clothing—Jenny and Jilly.

Thanks to the Royals for 1985 and the Hawks for 1988. Cheers to the gang from the summer of '85. Thanks to Star Trek and Jimmy Buffet.

Thanks to Sebago. Boo-hiss to the strip search squad at the Canadian border—twice.

Thanks to Lawrence, the mother-city. 10 years filled with so many memories I can't remember them all.

And of course to the Henry L. Doherty Fellowship and the Ward Graham Summer Fellowship for financial support.

How many graduate students does it take to screw in a lightbulb? Just one. But it'll take him 6 years.

List of Figures

List of Tables

Nomenclature

Symbol Description

$\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{a},\mathbf{b}$	Constants, calculation matrices/arrays, or vectors
A,B,C,D	Miller vapor pressure equation coefficients. Also, polynomial coefficients for enthalpy model.
C	Number of components in distillation operation
D	Distillate molar flowrate
$\boldsymbol{d}$	Vector of design variables (e.g., physical constants, column diameter, number of trays, heat exchanger area, utility temperatures), any dimension
F_j	Total molar flowrate of feedstream to distillation stage j
f_{ij}	Molar flowrate of component i in feedstream to distillation stage j
$\bar{f}_{ij}$	Mean value of the feedrate of component i to distillation stage j
F_P	Poynting correction factor
f	Arbitrary function
$\mathbf{g}$	Vector of process inequality constraints (design specifications, physical operating limits, etc), dimension m
g'	Active inequality constraint, i.e., most critical constraint, constraint with largest residual value
g_i	i'th inequality constraint, constraint of current interest
$\mathbf{H}$	Enthalpy, Btu/lbm
$\mathbf{h}$	Vector of steady-state process equations, (mass and energy balances, thermodynamics, kinetic constraints, etc), dimension n
i,j,k	Arbitrary counter variables
$\mathbf{K}$	K-value
L_j	Total liquid molar flowrate on distillation stage j
l_{ij}	Liquid molar flowrate of component i on distillation stage j
M	Total number of Monte Carlo simulations
m	Number of process inequality constraint equations - dimension of $\mathbf{g}$
N	Number of distillation stages

n	Number of steady-state process equations - dimension of $\mathbf{h}$. Also number of state variables $\boldsymbol{x}$
P	Pressure
P_i^{sat}	Saturated vapor pressure for component i, kPA
p	Number of uncertain parameters - dimension of $\boldsymbol{\theta}$
Q_j	Heat addition/removal on distillation stage j
q	Number of control variables - dimension of $\boldsymbol{z}$
$\mathfrak{R}$	Design reliability
$\mathfrak{R}_\Delta$	Net or final error in the design reliability estimate
$\mathfrak{R}_+$	Probability of positive error in reliability estimate (overestimation of reliability)
$\mathfrak{R}_-$	Probability of negative error in reliability estimate (underestimation of reliability)
$\mathbf{R}$	Space of $\boldsymbol{\theta}$ in which the constraints are all satisfied, i.e., the region inside the constraint boundary.
$\boldsymbol{R}$	Ideal gas constant
$\mathbf{S}$	Set of possible values of $\boldsymbol{\theta}$. Only used for min/max bounds.
$\mathbf{T}$	Set of possible control variable settings of $\boldsymbol{z}$
T	Temperature
U_j	Flowrate of liquid drawstream on distillation stage j
$\bigcup$	Design Unreliability
$\bigcup_j$	Probability that the j'th inequality constraint is violated
u	Standard normal random variable generated for the mean square successive difference test for randomness
$\boldsymbol{V}^*$	Characteristic volume (Hankinson-Thomson correlation)
$\boldsymbol{V}_i$	Liquid molar volume of component i (Hankinson-Thomson correlation)
V_j	Total vapor molar flowrate on distillation stage j
V^{Flood}	Vapor floodpoint for distillation column
v_{ij}	Vapor molar flowrate of component i on distillation stage j
W_j	Flowrate of vapor drawstream on distillation stage j
$\boldsymbol{x}$	Vector of state variables (temperature, pressure, composition), dimension n
$\hat{\boldsymbol{x}}$	Value of $\boldsymbol{x}$ at point $\hat{\boldsymbol{\theta}}$ on constraint boundary

$\hat{\boldsymbol{x}}^{\boldsymbol{i}}$	Value of $\boldsymbol{x}$ at point $\hat{\boldsymbol{\theta}}^{\boldsymbol{i}}$ on constraint boundary
$\boldsymbol{x}$	Liquid mole fraction
Y	Number of Monte Carlo simulation points inside constraint boundary
$\boldsymbol{y}$	Vapor mole fraction
Z_{α}	Standard Normal Quantile
z	Vector of control variables (equivalently degrees of freedom in process operation), dimension q
z^{*}	Optimum value of z

Greek Symbols

α	Level of significance (probability of type I error)
γ	Activity coefficient for liquid phase from liquid solution model
δ	One-dimensional variable to be solved in the line-search equation
$\hat{\delta}$	Solution of δ in the line-search equation
ε	Convergence tolerance
ε_{δ_1}	Nested-loop line-search convergence tolerance, type 1
ε_{δ_2}	Nested-loop line-search convergence tolerance, type 2
$\in$	Required accuracy in Chebyshev inequality
η	Tray efficiency
$\boldsymbol{\theta}$	Vector of process uncertainty parameters, dimension p
$\boldsymbol{\theta}^{\boldsymbol{S}}$	Vector of probabilistic (**S**tochastic) parameters, subset of $\boldsymbol{\theta}$
$\boldsymbol{\theta}^{\boldsymbol{F}}$	Vector of possibilistic (**F**uzzy) parameters, subset of $\boldsymbol{\theta}$
$\hat{\boldsymbol{\theta}}$	Point on constraint boundary
$\hat{\boldsymbol{\theta}}^{\boldsymbol{i}}$	Point i on constraint boundary. Also, orthogonal boundary point of inequality constraint i
λ_{ij}	Coefficients for Modified Wilson solution model - database uncertainty parameters
$\bar{\lambda}_{ij}$	Transformed parameter space of λ_{ij}
Λ_{ij}	Intermediate value of Modified Wilson solution model
ρ	p-dimensional radial direction vector from focalpoint to boundary, used in line-search equation, dimension [p·1]

$\hat{\phi}$	Fugacity coefficient for vapor phase from vapor solution model
ϕ_i^s	Fugacity coefficient of pure component i at T and P_i^{sat} from vapor solution model
ω	Acentric factor

Subscripts

Btm	Bottom
C	Critical property
D	Distillation overhead draw stream
Dist	Distillation
EB	Energy Balance
F	Feed stream
i	i'th value (also component number)
j	j'th value (also distillation stage number)
k	k'th value
L	Liquid stream
Max	Maximum value
MB	Material Balance
Min	Minimum value
n	n'th distillation stage (reboiler)
PE	Phase Equilibria
R	Reduced conditions
Top	Top
V	Vapor stream
+	Positive
–	Negative

Superscripts

abs	Absolute
BA	Boundary-approximation
Best	Best-estimate

D	Dummy (or slack) variable
Des	Nominal or design value
0,k,k+1	Iteration Number (0 is initial, k is current, k+1 is next step)
Max	Maximum value
Min	Minimum value
Opt	Optimistic
Pess	Pessimistic
Sim	Simulated value
pct	Percentage
T	Matrix Transpose
True	True value
-1	Matrix Inverse

Mathematical Functions

$\|\mathbf{A}\|$	Vector-Norm of vector $\mathbf{A}$, i.e., $= \left(\sum_{i}^{n} A_i^2 \right)$
$\|\cdot\|$	Absolute value
$cos^{-1}( )$	Inverse cosine function
$exp( )$	Exponential function
$int( )$	Greatest integer not greater than the argument
$ln( )$	Logarithm function
$\underset{i}{Max}\{A_i\}$	Maximum value of the i values of vector $\mathbf{A}$
$\underset{i}{Min}\{A_i\}$	Minimum value of the i values of vector $\mathbf{A}$
$\sum_{i=a}^{b}$	Sum from i=a to b
$\Delta \boldsymbol{x}$	Step change in x
∇	Gradient function
$\equiv$	"Is defined as"

Statistical Functions

$b\{a,b\}$	Binomial distribution with number of random samples (a) and probability (b)
$\mathbf{N}\{a,b\}$	Normal distribution with mean (a) and variance (b)
$\Pr\{\cdot\}$	Statistical likelihood (Probability) of $\{\cdot\}$ occurring
$\text{Poss}\{\cdot\}$	Statistical likelihood (Possibility) of $\{\cdot\}$ occurring
σ^2	Variance
$\mathbf{U}\{a,b\}$	Unform Probability distribution between (a) and (b)
$\chi^2\{\cdot\}$	Chi-Square Statistic

I. Introduction

Designing with uncertainty is to engineering as the mad uncle chained up in a dungeon is to the cursed medieval family—a dark secret ever present but never mentioned.

The analogy is only slightly exaggerated. While all engineers know their knowledge is limited, one might suggest that design with uncertainty is a neglected research area because engineers are scared of what they might find. It is always unsettling to examine one's own foundations too closely. Often one discovers quicksand.

> It is unlikely that we shall find satisfaction in looking too deeply into the foundations of our activities…often the more we look for the rationale behind what we are doing…the more remote it becomes until eventually we are driven to the conclusion that there is none (Gaines, 1984).

The antidote to process uncertainty is the addition of safety factors. However, selection methods are expensive, inexact and based primarily on rule-of-thumb experience. Worse, the design might *still* fail, and sub-optimal and indiscriminate oversizing of equipment can actually increase the likelihood of failure (Grossmann and Morari, 1984).

Increasing economic pressures require a more rigorous approach. Taking all uncertainties into account, find an 'optimum' design, i.e., one which maximizes the process reliability (odds of success) at minimum cost. Unfortunately, these goals are in opposition—improving reliability usually requires larger safety factors and greater capital cost. So the final design is a compromise.

These compromises must be selected by the designer. Therefore, he or she needs a computational tool to identify the critical uncertainties, evaluate the risks, and calculate the design reliability. With such information, the designer can increase safety factors for the critically undersized equipment, and reduce safety factors for the unnecessarily oversized equipment. This should lead to a near-minimum cost design with acceptable risk.

Early methods, e.g., Grossmann and Morari (1984), required all uncertainties to be described by min/max bounds. The problem was then to find the 'worst case' and check it. Unfortunately, the worst case is not always obvious, and even when successfully applied, the result is only a weak yes/no answer—either the design is guaranteed to work, or it *might* fail. The likelihood of failure is unknown. Also, min/max bounds are often difficult to quantify, and the calculation procedure must be repeated if any bound is changed.

Kubic and Stein (1988) extended the statistical theory to allow uncertainty descriptions using classical probability statistics and/or the more recent possibility or "fuzzy" statistics (min/max bounds are a subset of possibility statistics). Although undeveloped in chemical

engineering, possibility and probability (hybrid) statistics provide power and flexibility for describing uncertainties. Unfortunately, major computational limitations remain.

For probability statistics, there are two computational methods: (1) Conventional Monte Carlo, and (2) the Kubic and Stein (1988) method. Conventional Monte Carlo simulation is very simple and requires no process simulator modification. However, it is slow and imprecise. For example, 100 process simulations can only estimate a 90% reliability to within (-7.6%, +5.1%) at 95% confidence. After 1,000 simulations, the confidence interval is still (–2.0%, +1.8%). Even if ±2% imprecision is sufficient, the effort may be prohibitively expensive—if a single simulation takes only 5 minutes, 1,000 simulations will require 3.5 days.

The Kubic and Stein (1988) method is much faster. They calculate the statistical likelihood of violating each design constraint separately, a relatively easy task. From this, upper and lower bounds are found for the overall design reliability. However, these bounds are sometimes unacceptably wide, and certain mathematical simplifications make the accuracy questionable.

Recently, Straub and Grossmann (1990) and Pistikopoulos and Mazzuchi (1990) developed different methods for probability statistics. However, these are restricted to linear models—an unacceptable limitation for process design.

For possibility statistics, the solution requires a constrained optimization procedure. Kubic and Stein (1988) developed a reasonable, but fairly expensive, computational method.

For the most general case of hybrid statistics, Kubic and Stein (1988) used an extremely expensive variation of Monte Carlo. The possibility calculation (requiring several converged process simulations) must be performed *inside* the Monte Carlo loop. This is simply too expensive to be of practical use.

Finally, there are always uncertainties in the estimated uncertainties. Therefore, a complete study requires sensitivity analyses. The current methods require recalculation for each change, which is very expensive.

This thesis proposes an alternative calculation procedure that removes these barriers. The new procedure is accurate and precise and orders of magnitude faster than before. Sensitivity analyses can also be performed for insignificant cost.

Figure I-1 shows a picture of the fundamental mathematical problem. The x and y axes represent the various uncertainties or parameters. For some values of these parameters, the design succeeds (internal shaded region). For all other values, the design fails. The constraint boundary separates the two regions.

For probability statistics, the design reliability is the integral of the probability distribution throughout the region inside the constraint boundary. In general, this is too difficult to

calculate analytically. Conventional Monte Carlo generates a number of random points according to the probability distribution, then rigorously simulates each point to see if it is inside or outside the constraint boundary. The fraction of points inside the boundary is a statistical estimator of the design reliability.

The possibility statistical calculation optimizes a statistical function along the constraint boundary and requires several process simulations. The hybrid statistical calculation performs this function optimization inside a Monte Carlo loop.

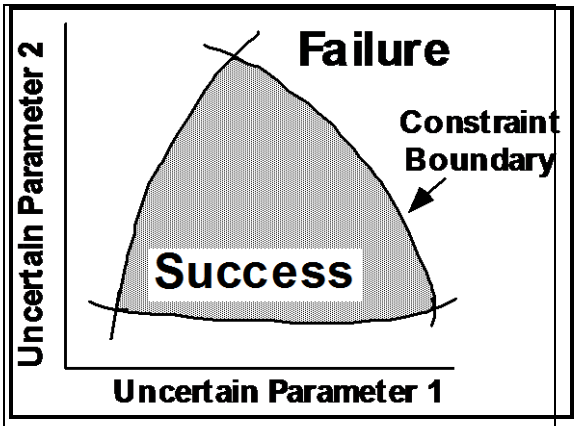


Figure I-1
Schematic of the Constraint Boundary

The mathematics are simple for all statistical cases, but require repetitive loops. The critical computational bottleneck is that the process simulations are *inside* these loops. That is, we have the algorithm shown in Table I-1:

Table I-1
Comparison of Existing and New Algorithms to Calculate Reliability

Existing Approach	New Approach
1) Statistical Calculation Loop a) Do Process Simulation End Loop	1) Map the Constraint Boundary 2) Perform the Statistical Computations

To separate the statistical calculations from the process simulations requires knowledge of the constraint boundary. The key is to separate the problem into two parts: (1) Map the constraint boundary. (2) Then perform the statistical computations.

Since the statistical calculation loop (step 2) does not require process simulation, it becomes trivial to perform and thus the critical bottleneck is removed. On the other hand, mapping the constraint boundary requires extensive process simulation effort which is slow and may create a new bottleneck. Luckily, this is not the case as this research will show.

Also, two additional statistical capabilities become possible: First, because the statistical calculation loop is many orders of magnitude faster than before, the imprecision of Monte Carlo integration can be made negligible. For example, a 100,000 point Conventional Monte Carlo integration can be done in minutes, reducing the uncertainty in the reliability estimate to a fraction of a percent. Second, sensitivity analyses simply repeat the statistical calculations and require no additional process simulations.

Mapping the constraint boundary requires development of a computationally efficient mathematical model of the constraint boundary. This can be done by finding several individual points on the constraint boundary, then interpolating between them in some way—surface approximation mathematics. However, since interpolation is unavoidably approximate, so is the design reliability estimate.

The best approach is to use an iterative algorithm as outlined in Figure I-2. Start with a rough, initial boundary approximation, then find additional boundary points as needed to improve the accuracy. By concentrating the boundary points in the most statistically significant boundary locations, computational effort is minimized. Note that the designer can identify and correct a poor design at the earliest possible stage.

This work develops and tests the boundary-approximation procedure. First, the statistical rigor is proved using geometrical problems. Then, the procedure is compared to Conventional Monte Carlo using binary and ternary component distillation as the test basis. This research is limited to probability statistics only. However, the procedure is extendible to hybrid statistics.

This thesis will show that the boundary-approximation procedure is easily superior to all other current methods for estimating design reliability. While there are remaining limitations, these can be eliminated by further development.

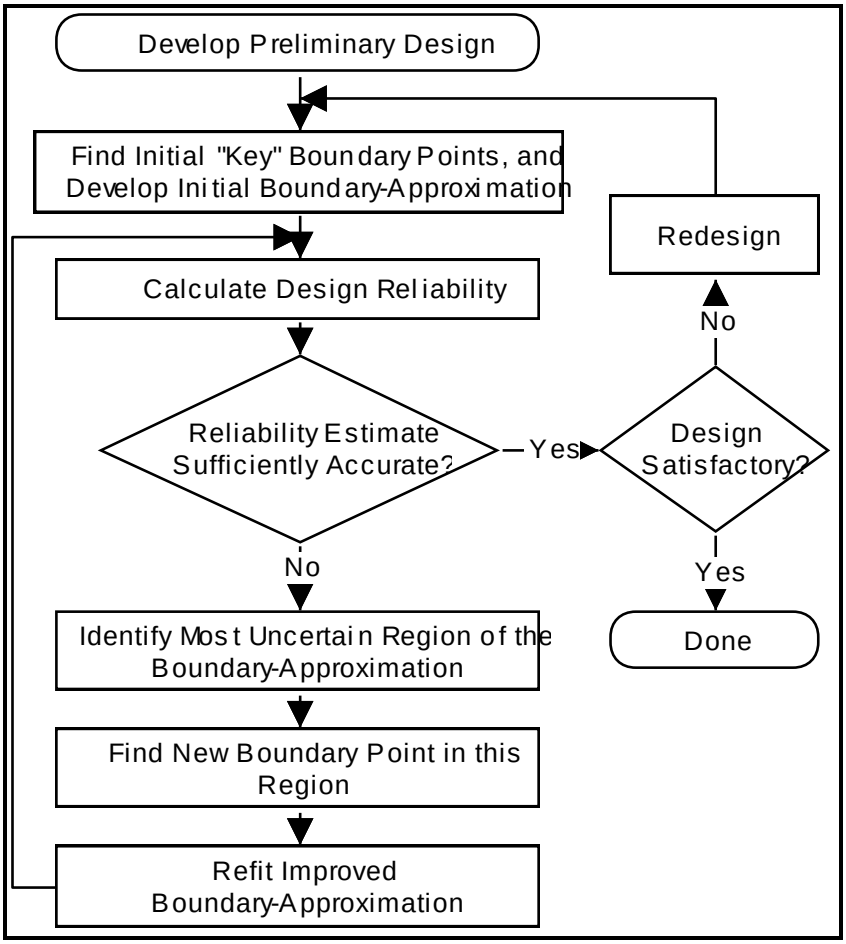


Figure I-2
Flowchart of Design Evaluation and Reliability Calculation

II. Process Design Reliability Development

This section develops the background for the design reliability calculation procedure. First, the fundamental design problem is developed (II.A). Then, the basic mathematical picture is presented (II.B). Next, the statistical representation of process uncertainties is discussed (II.C) Finally, currently available design reliability calculation procedures are reviewed (II.D).

II.A. Process Design Objective

This section formulates the fundamental design reliability problem: Given a proposed design and a range of possible conditions and/or uncertainties, calculate the design reliability. Other topics such as economic optimization and process synthesis are related, but scoped out of further consideration.

A process design always includes safety factors to ensure satisfactory operation over a range of operating conditions (e.g., feed flowrates), and to allow for design model uncertainties (e.g., thermodynamic database and distillation tray efficiency). Because safety factors are expensive, their selection should be an optimized compromise between risk and cost. This requires a statistical analysis of the uncertainties involved.

Three alternative problem formulations are (Grossmann and Morari, 1984):

1) Given a proposed design, calculate its reliability.

2) Develop a process design with a specified reliability.

3) Design a process with optimum reliability.

Note: Reliability is defined as the statistical likelihood that the design will succeed.

The first approach is the most basic of the three: For any proposed design and specified range of possible conditions and/or uncertainties, calculate the likelihood that the design will succeed.

The second approach is a straightforward extension of the first: Find the minimum-cost design that guarantees a certain degree of reliability, given a specified range of conditions and uncertainties. This requires an extra level of iteration above the first.

The third approach requires quantification of ‘optimum’ reliability which is difficult. Clearly this requires engineering judgment and some trade-off between economics and risk. In general, increasing the design reliability requires larger safety factors and increased capital cost. This leads to a range of possible solutions—a Pareto-optimum (Nishida and Ohtake, 1986; Diaz, 1987), see Figure II.A-1. The designer must choose the most acceptable point on this curve.

In practice, a designer will use an interactive approach: (1) Develop a preliminary design. (2) Calculate the reliability, identify the critical uncertainties, and determine which equipment is undersized and which is oversized. (3) Modify the design and repeat the analysis, continuing until satisfied.

All these situations can be handled with one fundamental calculation tool: Calculate the design reliability given a proposed design and its uncertainties.

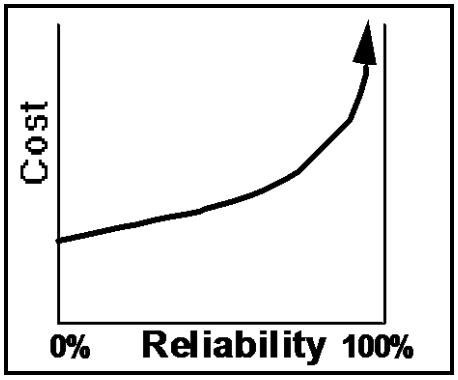


Figure II.A-1
Trade-Off Between Cost and Reliability

Simultaneous Process Design and Optimization

One of the main areas of process design research has been *simultaneous* process design and economic optimization (Biegler, 1984, 1988; Buzzi-Ferraris and Tronconi, 1985; Chen and Stadherr, 1985a,b,c; Kumar and Lucia, 1987). These methods are concerned with optimizing an economic objective function, e.g., capital cost + two years operating expenses. However, in practice safety factors must still be added afterwards, which increases the capital cost and invalidates the economic optimum.

Ganesh and Biegler (1987) developed an efficient method to calculate first-order sensitivity of the objective function to variations in the parameters and process model (see also Buzzi-Ferraris and Tronconi, 1985). However, while this can help the designer ensure that none of the uncertainties are economically sensitive, the design is not necessarily optimum.

Instead of optimizing a single economic objective function, a better design goal is to optimize the economics over the range of possible operating conditions. Here, optimum could be defined as an average expected profit (Wen and Chang, 1968), or a minimized worst case (Nishida et al., 1974).

Another argument against optimizing a single economic objective function is that statistical design should lead to nearly optimal economic designs anyway: (1) The designer can identify the critical equipment and provide the necessary safety factors. (2) The designer can tighten up the design where the equipment is unnecessarily oversized. These two steps minimize the capital costs while providing high reliability.

Perhaps the economics are not as critical as they seem. The point can be made that a process would not be built unless it is profitable and/or necessary, therefore the focus should

be less on rigorously optimizing the economics and more on ensuring reliability and operating flexibility in the future. As stated by Fisher et al. (1985a):

> The procedure generally followed in industry, however, is to estimate the profitability of a process operating at some base-case conditions, using nominal values for all model parameters. Empirical overdesign factors are then applied to each piece of process equipment to ensure operability. If the process still appears to be profitable, more rigorous design and economic models are used. The effect of parameter uncertainty is not then considered until the most rigorous design models are developed and the process profitability is assured.

Another interesting point was made by Rippin (1992):

> (For) a plant operating in a changing and uncertain environment…it will surely not be appropriate to design such a plant to a single reference performance at minimum cost. (I remember a senior engineer at a symposium…more than 20 years ago—when optimization was just becoming fashionable—saying something like: "When we built a plant in the old days, we knew conditions would always change after the plant was built, so we put in a bottleneck here and a bottleneck there, so that when conditions did change the bottlenecks could be relieved one after another, gradually modifying the performance without too much expense. These new 'optimized' plants are no use at all; they bottleneck all over the place, all at the same time and you cannot do a thing with them.")

While I think the senior engineer was giving himself a little too much credit for forethought—without the analysis tools, design bottlenecks are more likely to be accidental than strategic—his final point is valid: A design optimized for a single economic function might be very poor in other areas, such as operating flexibility. On the other hand, a statistical design can provide an "economically appropriate amount of flexibility taking account of the *value* of the flexibility in the anticipated environment, and the *cost* of providing it" (Rippin, 1992). It can also guarantee that a design *will* work in those all too common situations when the uncertainties are highly significant.

Simultaneous Process Synthesis and Optimization

Another related topic is simultaneous process *synthesis* and optimization, where the process simulator is not only expected to optimize the economics, but must also develop the best flowsheet—optimum processing sequence and heat integration. While this is a very difficult task in general, useful techniques do exist and it is a major research area (Nishida et al., 1974; Floudas et al., 1986; Papoulias and Grossmann, 1983; Fisher et al., 1985b; Duran and Grossmann, 1986; Kocis and Grossmann, 1987; Árva and Czukás, 1987, 1988). Process synthesis can be incorporated in the future. However, it will not be considered here.

Process Design Including System Dynamics

Another important topic is process design taking system dynamics into account, i.e., optimizing controllability and resiliency of the process (Grossmann and Morari, 1984). Since it is difficult and expensive to correct for problematic dynamics in the late stages of design, process control engineers recommend that this be considered in the earliest possible stages (Brignole et al., 1985; Moore, 1985; Fisher et al., 1988abc; Floudas, 1987; Palazoglu and Arkun, 1986, 1987; Skogestad and Morari, 1987c). Only steady-state will be studied in this work, however.

II.B. Mathematical Formulation of Design Reliability Problem

This section develops the fundamental mathematics of the design reliability calculation problem. Although the problem solution is complicated, a very simple geometric visualization is possible.

A chemical process flowsheet consists of process operation equipment connected by flowstreams. This equipment includes distillation columns, reactors, heat exchangers, pumps etc. The engineer has a design objective (e.g., perform a chemical separation), and it is his or her job to develop the flowsheet and specify the equipment sizes. To do this, a mathematical process model must be developed and solved.

The process model consists of the material and energy balances, thermodynamic and kinetic equations, etc. These can be represented by a vector (or set) of constraint equations $\mathbf{h}$. The unknowns to be solved for are the stream variables (temperature, pressure, composition), also known as "state" variables. These can be represented by a vector $\boldsymbol{x}$. So the process model can be written:

$$\mathbf{h}(\boldsymbol{x}) = \mathbf{0} \qquad \text{(dimension n)} \qquad \text{Eq II.B-1}$$

Eq II.B-1 can be written for individual process equipment or the entire flowsheet. Solution methods are generally iterative.

Additional design specifications or constraints must also be met. For instance, a distillation column performs a chemical separation where the product streams must meet minimum purity specifications. Other examples include flooding limits in the column, temperature and pressures within acceptable equipment operating ranges, maximum pumping rates, etc. All these constraints can be written as inequality equations, denoted by vector $\mathbf{g}$:

$$\mathbf{g}(\boldsymbol{x}) \le \mathbf{0} \qquad \text{(dimension m)} \qquad \text{Eq II.B-2}$$

where by convention the constraint is satisfied when $\mathbf{g}$ is negative and violated when $\mathbf{g}$ is positive.

Then there are the design variables $\boldsymbol{d}$, set by the designer. These describe the physical dimensions of the equipment and fixed plant operating conditions, such as steam and cooling water temperatures.

Most process operations have control variables or operating degrees of freedom, denoted by vector z. These are controlled by the operator and can be varied within certain limits to optimize process operation. For example, distillation control variables might consist of column pressure, reflux ratio, and vapor boilup rate. These are adjusted to maximize the separation for minimum energy usage.

Finally, the process model is uncertain due to physical property database uncertainties and imprecise knowledge of the equipment operation. For example, the distillation model requires enthalpies and K-values which must be estimated from empirical equations fitted to limited and imprecise data. Tray efficiencies, flooding points, and heat transfer coefficients are uncertain. Also, the design problem itself might be uncertain, with imprecisely known feed stream compositions and economic factors.

These process uncertainties are represented by *parameters*, denoted by θ. These are constants which instead of being a single value, are described by a range of values such as min/max bounds or statistical distributions. For instance, we might estimate tray efficiency to be somewhere between 40% and 60%.

Now z may be varied to compensate for poorly estimated θ. That is, z provides process flexibility. But z has a fixed range, so some operating limit exists.

The design objective, therefore, is to design the process (choose $\boldsymbol{d}$) such that the constraints are met ($\mathbf{h}=\mathbf{0}$, $\mathbf{g}\leq\mathbf{0}$), given the possible values of θ.

The traditional approach is to assume best and/or worst case estimates for θ, solve the process model for $\boldsymbol{x}$, then size the equipment including safety factors. The hope is that even if the estimated θ is incorrect, the process will still work. However, a better approach is to quantify the risks involved. Specifically, for the given design, what is the design reliability, i.e., the statistical likelihood that it succeeds? To summarize:

- For a given design $\boldsymbol{d}$ chosen by the designer…
- and for the possible values of θ…
- and optimizing z (denoted z^*) over its operating range $z \in \mathbf{T}$ (optimum z^* occurs where $\mathbf{g}$ is minimized):

  Calculate $\mathfrak{R}$ ($\mathfrak{R}$ = Design Reliability) where:

  $$\mathbf{h}\left(\boldsymbol{d}, z^*, \boldsymbol{x}, \theta\right)=\mathbf{0} \qquad \text{Eq II.B-3}$$

  and

$$\Re \;=\; \Pr\left\{\mathbf{g}\left(\boldsymbol{d}, z^{*}, \boldsymbol{x}, \boldsymbol{\theta}\right) \le \mathbf{0}\right\} \qquad \text{Eq II.B-4}$$

Pr{ · } denotes the statistical likelihood

While this is a difficult problem to solve, there is a simple visualization. First, to simplify note that for any value of $\boldsymbol{\theta}$, there exists an optimum z^{*}. So we can eliminate z^{*} as an independent variable:

$$z^{*} \;=\; z^{*}[\boldsymbol{d}, \boldsymbol{x}, \boldsymbol{\theta}] \qquad \text{Eq II.B-5}$$

which allows us to write Eq II.B-3 and Eq II.B-4 as:

$$\mathbf{h}\left(\boldsymbol{d}, z^{*}[\boldsymbol{d}, \boldsymbol{x}, \boldsymbol{\theta}], \boldsymbol{x}, \boldsymbol{\theta}\right) = \mathbf{0} \;\Rightarrow\; \mathbf{h}(\boldsymbol{d}, \boldsymbol{x}, \boldsymbol{\theta}) = \mathbf{0} \qquad \text{Eq II.B-6}$$

and

$$\Re = \Pr\left\{\mathbf{g}\left(\boldsymbol{d}, z^{*}[\boldsymbol{d}, \boldsymbol{x}, \boldsymbol{\theta}], \boldsymbol{x}, \boldsymbol{\theta}\right) \le \mathbf{0}\right\} \;\Rightarrow\; \Re = \Pr\left\{\mathbf{g}(\boldsymbol{d}, \boldsymbol{x}, \boldsymbol{\theta}) \le \mathbf{0}\right\} \qquad \text{Eq II.B-7}$$

Likewise, we can use $\mathbf{h}$ to solve for and eliminate $\boldsymbol{x}$:

$$\mathbf{h}(\boldsymbol{d}, \boldsymbol{x}, \boldsymbol{\theta}) = \mathbf{0} \;\Rightarrow\; \boldsymbol{x} = \boldsymbol{x}[\boldsymbol{d}, \boldsymbol{\theta}], \text{ given } \mathbf{h} \qquad \text{Eq II.B-8}$$

which we can substitute into $\mathbf{g}$ to simplify:

$$\mathbf{g}\left(\boldsymbol{d}, \boldsymbol{x}[\boldsymbol{d}, \boldsymbol{\theta}], \boldsymbol{\theta}\right) \;\Rightarrow\; \mathbf{g}(\boldsymbol{d}, \boldsymbol{\theta}) \qquad \text{Eq II.B-9}$$

and the design reliability is simply:

$$\Re = \Pr\left\{\mathbf{g}(\boldsymbol{d}, \boldsymbol{\theta}) \le \mathbf{0}\right\} \qquad \text{Eq II.B-10}$$

Since $\mathbf{d}$ is a vector of fixed constants set by the designer, we can reduce Eq II.B-10 to:

$$\Re = \Pr\left\{\mathbf{g}(\boldsymbol{\theta}) \le \mathbf{0}\right\} \qquad \text{Eq II.B-11}$$

and with the final simplification of expanding the m-dimensional vector $\mathbf{g}$:

$$\Re = \Pr\left\{\left[\underset{i=1\ldots m}{Max}\, \mathrm{g}_{\mathrm{i}}(\boldsymbol{\theta})\right] \le 0\right\} \qquad \text{Eq II.B-12}$$

Eq II.B-12 shows that the reliability $\Re$ is a function only of $\boldsymbol{\theta}$, and thus may be mapped in p-space. Each dimension represents an uncertain parameter $\boldsymbol{\theta}_{\mathrm{j}}$ for $\mathrm{j}=1\ldots\mathrm{p}$. For illustration, we use p=2 and the xy-plane, see Figure II.B-1.

Each inequality constraint g_i is an independent function in p-space. A boundary is created at the equality, i.e., $g_i(\theta)=0$. On one side of the boundary, that particular constraint is satisfied. On the other side, the constraint is violated. Figure II.B-1 shows boundaries for three inequality constraints, which together determine the constraint boundary. The 'internal intersection' of the constraints forms an enclosed volume. Inside this volume, all the constraints are met and the design succeeds (the shaded region in Figure II.B-1). Outside this volume, at least one constraint is not met, and the design fails.

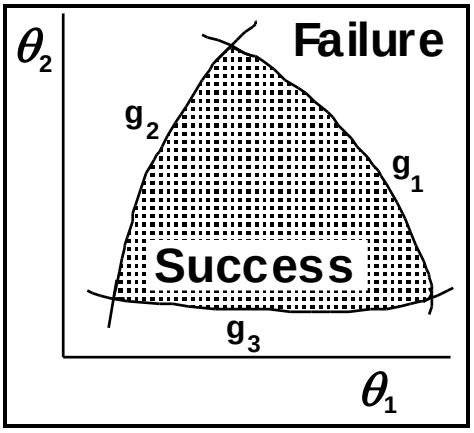


Figure II.B-1
Design Region Showing Constraint Boundary

Therefore, the design reliability is equal to the statistical likelihood that the true value of θ is inside the constraint boundary. This is the basic mathematical problem.

This picture will be returned to later. Next we present the statistical representation of the uncertain parameters.

II.C. Statistical Description of Process Uncertainties

This section discusses the mathematical techniques for describing process uncertainties. Mathematical detail is kept to a minimum as this research does *not* develop these techniques. This section is only intended to show: (1) the potential of these descriptive techniques, and (2) the capabilities that the reliability calculation procedure must provide.
Probability statistics should be familiar to most engineers. However, possibility statistics is a complementary technique that is untested in chemical engineering.

A design reliability estimate can only be as good as one's knowledge of the uncertainties. Some process design uncertainties are well known and easily quantified, e.g., phase equilibrium when the thermodynamics models are adequate and sufficient data are available. Other uncertainties are poorly known and extremely difficult to describe, e.g., feed composition variations in future operation.

Especially for the latter situation, the standard mathematical descriptions of uncertainty are inadequate. While promising new techniques exist, they are barely in the research phase. This research does *not* develop these techniques. However, they are briefly described in the following sections. This serves to: (1) Show there is hope for satisfactory description of the process uncertainties in the future, and (2) Identify the types of uncertainties that the reliability calculation procedure must ultimately provide for.

II.C.1. Min/Max Bounds

This is the simplest form of uncertainty representation, and has been used most often in the historical development of process design reliability. However, it is a very weak description and has been superseded by possibility statistics.

The simplest type of uncertainty representation is minimum and maximum bounds for the value of the parameter. This type has been used most often in the historical development of process design reliability procedures. However, there are two major limitations with this representation:

1) It is often difficult or impossible to obtain precise values for these min/max bounds.
2) Any additional information, such as a 'best-estimate' or most-likely value, cannot be used.

Min/max bounds can be represented using the more general possibility statistics, and are now obsolete.

II.C.2. Probability Statistics

This is the most powerful mathematical representation of uncertainty. However, it requires a significant amount of information to be valid.

Probability statistics should be familiar to most engineers and is the most powerful uncertainty representation. However, because the statistics are so powerful, a significant amount of information is required for accurate representation. If the information is

unavailable, one must rely on unjustified and potentially incorrect assumptions. Thus, this representation is sometimes *too* powerful and cannot adequately describe situations where the information is highly uncertain.

For instance, say a distillation tray efficiency is 'somewhere' between 40% and 60%. This is a min/max bound, but using probability statistics we might specify a uniform probability distribution as shown in Figure II.C.2-1. However, this implicitly describes more information than is actually known, for example that the tray efficiency has a 50% likelihood of being between 50% and 60%.

Even though probability statistics cannot correctly describe all situations, it is often used regardless, because no alternative existed. Users have simply had to hope that their analysis is "robust", i.e., that the results are nearly correct, even though the assumptions might not be.

Some specific situations in which probability statistics cannot cope are:

1) Min/max bounds are sometimes all that is known—and even these are uncertain.
2) Process simulation model errors are systematic, not random, which makes them very difficult to describe using probability distributions.
3) Empirical models must frequently be extrapolated into unknown regions, where it is generally impossible to quantify the uncertainties.

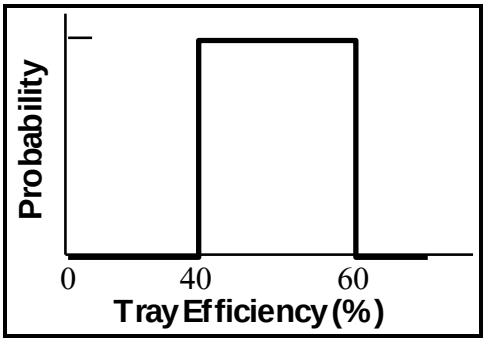


Figure II.C.2-1
Uniform Probability Distribution

Fortunately, possibility statistics provides a potential solution to these situations, as discussed below.

II.C.3. Possibility or "Fuzzy" Statistics

Possibility statistics is a new technique which fills the gap where probability statistics is inadequate. However, it is undeveloped. This section is intended only to show the potential, using a minimum of technical development.

Possibility or "fuzzy" statistics (Zadeh, 1968, 1978, 1988; Dubois and Prade, 1988) is more suitable than probability statistics when information is highly uncertain. It requires fewer assumptions and can represent less powerful information such as min/max bounds.

Possibility distributions are qualitative instead of quantitative. They measure *potential* of occurrence, whereas probability distributions measure *frequency* of occurrence. A possibility distribution is less precise and contains less information than a probability distribution.

Consequently, results will always be weaker than the equivalent probability analysis. However, this is not a disadvantage because possibility statistics are useful in precisely those situations where detailed information is unavailable.

It must be emphasized that possibility distributions are only mathematical tools for describing uncertainty, and there is some controversy over their theoretical 'meaning'. At the very least, possibility distributions accurately describe min/max bounds, as shown in Figure II.C.3-1 (graph on left). But it is not clear how to accurately describe additional information such as best-estimates. One could use a trapezoidal or triangular distribution, as shown in the middle and right graphs respectively. These distributions still correctly show that the actual value can be anywhere within the min/max bounds, but they also show that the true value is *more* possibly somewhere in the middle. Unfortunately, there does not yet appear to be a rigorous basis for selecting the best distribution (e.g., trapezoid, triangular or curved), and naturally the reliability estimate is a function of this distribution.

Hopefully, this difficulty will be overcome in the future. For the moment, possibility distributions are at least as good as min/max bounds, and with luck will prove superior.

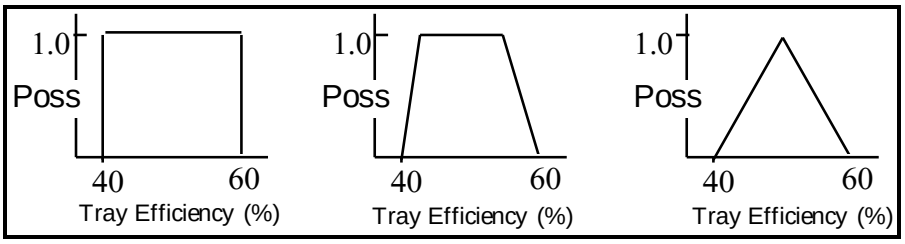


Figure II.C.3-1
Possibility Distributions with Same Min/Max Bounds but Progressively More Precise Information

II.C.4. Empirical Models and Extrapolation

This section discusses two particular chemical engineering problems: (1) Models are empirical, and therefore inherently 'wrong'. (2) Designers must frequently extrapolate these models into regions for which no data exist.

Rarely is a chemical engineering model theoretically correct. Often the models are simply empirical curve-fits of imprecise data. So not only are the model parameters uncertain, but so is the model.

Mathematically, one could define a 'dummy' parameter to represent the model uncertainty over and above that of the model parameters. We can write:

$$f(\theta) + \theta^{D} = 0 \qquad \text{Eq II.C.4-1}$$

where

f is an empirical model equation

θ is the empirical model parameter(s)

θ^D is the dummy parameter describing the 'leftover' uncertainty

θ^D is difficult to represent statistically, because the model errors are systematic, not random, i.e., the model is consistently low or high for particular values. So probability statistics might not be suitable. Possibility statistics can be applied two ways: (1) Regress θ (not including θ^D) using probability statistics as usual, then develop θ^D from the regression error residuals. (2) Estimate hybrid statistics for all θ including θ^D. The latter requires fuzzy regression (Tanaka et al., 1982; Yager, 1982; Darby-Dowman et al., 1988) and should be more rigorous, however the former is easier to implement.

A closely related problem is model *extrapolation*—a particularly difficult situation when the model is empirical. A current severe industrial example of this is supercritical fluid extraction, where the thermodynamic models are known to be inadequate and there are no data in the region of interest. How can one accurately quantify the error in extrapolating an empirical model into an unknown region?

I know of no way to do this. However, one can 'qualify' the potential error. Kubic and Stein (1986, 1987, 1988) developed an ingenious approach using possibility statistics, using the principle of "system similarity": If a process model works well for one system (where a "system" could refer to a particular chemical mixture at a certain temperature and pressure), it is likely to work well for "similar" systems (of close chemical composition, temperature and pressure).

If "similarity" can be quantified, it can be correlated with model error for known systems. Then the potential error can be estimated for a new system based on its similarity to the known systems. The key statistical development is that the final information is not a probability estimate of error, but an estimate of the uncertainty in the predictions. That is, along with the predicted value, we estimate the possibility of being correct and the possibility of being incorrect.

These possibilities are independent. A small extrapolation with a good model gives a high possibility of success and a low possibility of failure. A large extrapolation with the same model gives a high possibility of success *and* a high possibility of failure. Applying this approach to the supercritical fluid extraction example would give a low possibility of success and a high possibility of failure—if the knowledge base is inadequate, the analysis says so.

While this approach is promising, Kubic and Stein's work was only a first-pass attempt. The idea is completely untested and requires much more development. The key difficulty is quantifying similarity.

II.C.5. Uncertainty in the Uncertainty

This section warns that all statistical descriptive techniques are inherently uncertain themselves, and unavoidable assumptions are always made that might be incorrect. One should therefore analyze the potential impact of these assumptions. This can be accomplished by sensitivity analysis.

"As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality." Albert Einstein.

All uncertainties are themselves uncertain. Unavoidable assumptions are always made that might be incorrect. The shape or form of a statistical distribution is itself an assumption. Therefore, a complete analysis of the design reliability must consider what might happen if the assumptions are incorrect.

One way to do this is to perform a sensitivity analysis: First, do the calculations as normal. Then, change each assumption and redo the calculations. If the results are very different, the assumption is critical. To avoid this, the designer has two alternatives: (1) Increase safety factors to reduce sensitivity to the assumption. (2) Remove the assumption by improving the process model.

While sensitivity analysis is easy to perform, it becomes computationally expensive if there are many assumptions and each calculation is lengthy.

II.D. Available Methods to Calculate Design Reliability

This section reviews the available methods to calculate design reliability prior to this research. Most early work was limited to min/max bounded uncertainties (II.D.1). For probability uncertainties, Conventional Monte Carlo integration is a common but expensive approach (II.D.2.a). Kubic and Stein (1988) proposed a shortcut (II.D.2.b) which is inexpensive, but imprecise and potentially inaccurate. For hybrid (possibility+probability) uncertainties, Kubic and Stein (1988) developed the rigorous statistics (II.D.3 and II.D.4). But their calculation procedure is imprecise, potentially inaccurate, and computationally very expensive.

II.D.1. Min/Max Bounds

This section summarizes the historical development of design reliability calculation when all uncertainties have min/max bounds. This has been the main historical focus. However, calculations are expensive and the rewards are puny—either the design *will* work, or it *might* fail.

The historical work will not be reviewed in detail. For a comprehensive listing of early work, see Grossmann and Morari (1984).

Grossmann and coworkers (e.g., Grossmann and Sargent, 1978; Grossmann et al., 1983; Halemane and Grossmann, 1983; Grossmann and Morari, 1984; Swaney and Grossmann, 1985a,b; Grossmann and Floudas, 1987) have been the main historical force in process design reliability calculation. Grossmann and Sargent (1978) were the first to define the complete problem for min/max bounded uncertainties, but early solutions were prohibitively expensive. Work has progressed to the point that large-scale designs are do-able (Grossmann and Floudas, 1987), although to be realistic only the feasibility of these methods has been demonstrated, and there is still far to go for efficient implementation (Biegler, 1988).

These methods assume that the only information known about the parameters are min/max bounds. This is equivalent to a hyper-rectangle in the parameter space, see Figure II.D.1-1. Then, it is a matter of determining whether this region is completely enclosed by the constraint boundary (Figure II.D.1-1a), or if part of it overlaps the boundary (Figure II.D.1-1b). In the former case, the design is guaranteed to succeed. In the latter case, there is a potential for failure.

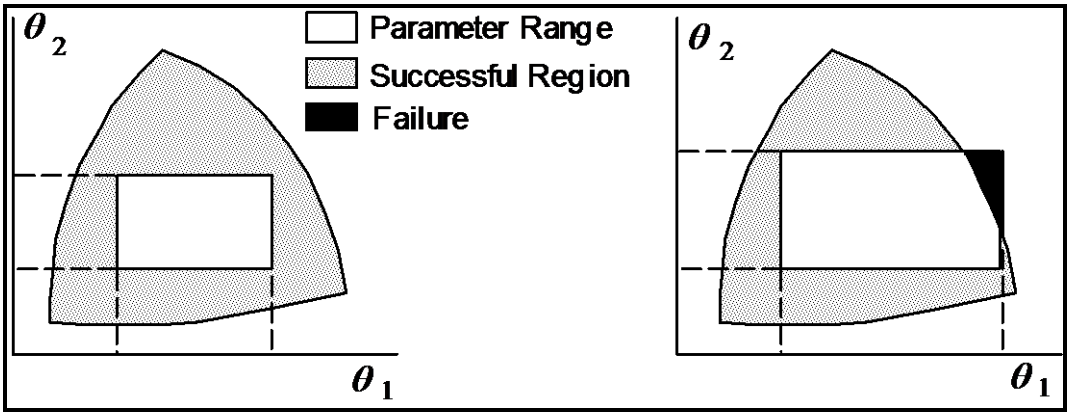


Figure II.D.1-1a
Reliable Design: All Parameter Values Satisfy Design Constraints

Figure II.D.1-1b
Unreliable Design: Some Parameter Values Violate Design Constraints

Mathematically, this can be formulated as follows: For the process design to be guaranteed, the set of inequality constraints $\mathbf{g}$ must be less than or equal to zero for all potential values of θ. This can be written as:

$$\underset{\theta \in S}{Max}\ \underset{z \in T}{Min}\ \underset{i=1\ldots m}{Max}\ g_i(d,z,\theta) \;=\; \begin{cases} \leq 0 & \text{Design guaranteed} \\ > 0 & \text{Design } \textit{might} \text{ fail} \end{cases} \qquad \text{Eq II.D.1-1}$$

where $\mathbf{S}$ is the set of all possible values of θ as defined by Eq II.D.1-2, and subscript i is the i'th inequality constraint equation.

$$\mathbf{S} \;=\; \left(\theta \mid \theta^{Min} \leq \theta \leq \theta^{Max}\right) \qquad \text{Eq II.D.1-2}$$

Eq II.D.1-1 can best be explained from inside-out: $\underset{i=1\ldots m}{Max}\, g_i(d,z,\theta)$ refers to the worst or 'active' constraint, i.e., the one with the highest value at the current conditions. This is minimized with respect to the control variables z over their range of values $\mathbf{T}$ so as to always obtain the best possible operation. Finally, this amount is calculated over the entire range of the parameter space $\mathbf{S}$, to find the worst case. After all this, if the amount is less than zero, then the process is guaranteed to work. Else, there exist some values of θ in $\mathbf{S}$ for which the process will fail.

The max-min-max problem is extremely difficult to solve, since it is essentially an infinite, non-differentiable, nonlinear programming, optimization problem. The problem is identifying which set of parameters is the worst case (e.g., the top-right vertex in Figure II.D.1-1b). Even assuming that the worst case is at one of the vertices, which is only true for the linear and convex cases, an exhaustive search may be prohibitively expensive in

practice, since there are 2^p vertices (p =number of uncertain parameters). Swaney and Grossmann (1985a,b) obtained some simplification, then Grossmann and Floudas (1987) noted that at any one time, only $q+1$ (q is the number of control variables z) of the inequality constraints are active, from which the optimum z^* can be solved directly from θ. This reduces the max-min-max problem to a single optimization in θ. The only difficulty remaining is to determine which set of constraints is active for the given θ.

For the linear constraints case, the problem reduces to a standard mixed-integer linear problem (MILP), for which many efficient solution methods exist. For nonlinear constraints, a mixed-integer nonlinear problem (MINLP) results, which can be reformulated as a set of nonlinear programming (NLP) cases. All possible sets of active constraints are first identified (>q in general, but usually a fairly small subset), then the calculations are performed for each possible combination.

It is not necessary to assume that the worst case is at one of the vertices, although this will be true with linear and convex constraints. And an exhaustive search is not required. For the nonlinear case, the optimal solution is guaranteed provided only that the nonlinear constraints are monotone in the feasible region.

Grossmann and Floudas (1987) say that computational requirements are reasonable and non-problematic even for large systems. However, even so there are several severe limitations:

1) Computations are time-consuming.
2) Only min/max bounds can be handled. Additional information, such as statistical distributions or 'best-estimate' values, cannot be used.
3) Min/max bounds might be uncertain themselves. Changing these bounds requires the calculation to be repeated.
4) Final knowledge is very limited. Either the design is guaranteed to work, or it *might* fail. There is no estimate of design reliability.

II.D.2. Probability Statistics

This section describes the calculation of design reliability for probability uncertainties. Mathematically, the design reliability is equal to the integral of the probability distribution throughout the volume inside the constraint boundary.

Although some early statistical design work in the 1960's and 1970's used probability statistics, no one tackled the complete problem. See Grossmann and Morari (1984) for a listing.

Geometrically the hyper-rectangle in Figure II.D.1-1 is replaced by a statistical distribution surface. Figure II.D.2-1 shows a representation in two dimensions, where both

parameters have a normal distribution. This can be visualized as a Gaussian bell-shaped dome 'above' the parameter space. The horizontal axes denote the parameter values θ^S. The vertical axis is the probability of this value occurring.

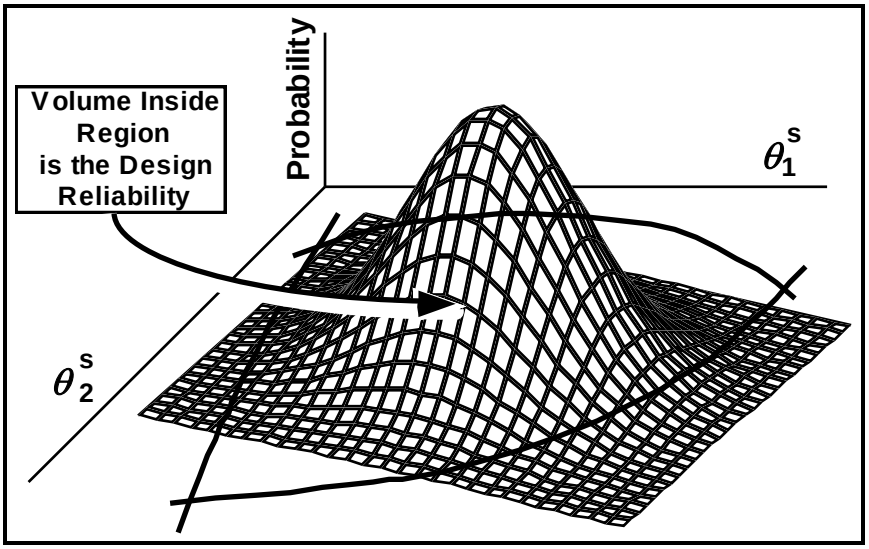


Figure II.D.2-1
Normal Probability Distribution in the Design Region

The reliability calculation problem becomes: What is the statistical likelihood that the true parameter values are inside the constraint boundary? Mathematically, this is equal to the integration of the probability distribution throughout the region enclosed by the constraint boundary:

$$\Re = \int_{\theta \in \mathbf{R}} \Pr\{\theta\} d\theta \qquad \text{Eq II.D.2-1a}$$

$$= \int \ldots \int \Pr\{\theta_1, \ldots, \theta_p\} \; d\theta_1 \ldots d\theta_p \qquad \text{Eq II.D.2-1b}$$

where

$\Re$ is the reliability $(0 \le \Re \le 1)$

$\mathbf{R}$ is the region inside the constraint boundary

If the constraint boundary is known exactly, this integration could be done analytically. However, this is not done in practice, due to the complex nature of the boundary which is a multi-dimensional, curved polyhedron, and the fact that some probability distributions cannot be integrated analytically.

II.D.2.a. Conventional Monte Carlo Solution Procedure

Conventional Monte Carlo is the traditional approach to estimating design reliability for probability distributions because it is very simple. Unfortunately, it is time-consuming and imprecise.

Conventional Monte Carlo integration (Rubinstein, 1981; Kalos and Whitlock, 1986) simply generates a set of random points according to the probability distribution, then performs a process simulation at each point to determine if it is inside or outside the boundary. The fraction of points inside the boundary is a statistical estimate of the reliability. This is shown schematically in Figure II.D.2.a-1. The basic algorithm is outlined in Table II.D.2.a-1.

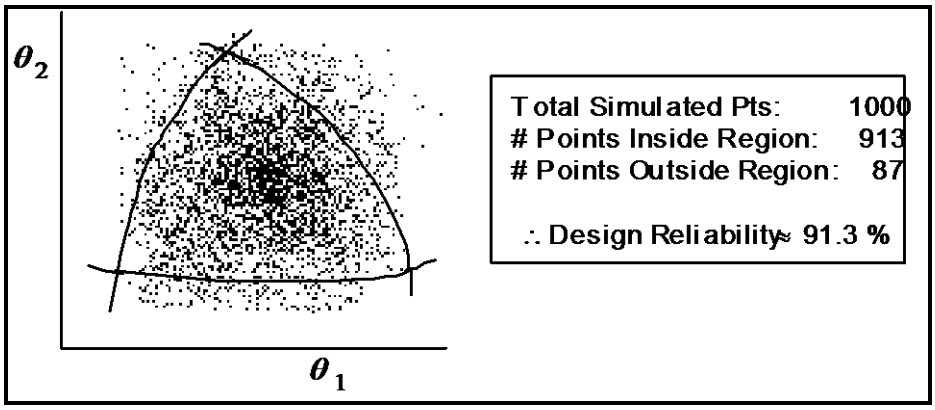


Figure II.D.2.a-1
Monte Carlo Integration to Estimate Design Reliability
(Probability Statistics)

Monte Carlo integration is straightforward for any number of parameter uncertainties and boundary nonlinearities. Consequently, it has been the technique of choice in most probability studies (Lashmet and Szczepanski, 1974; Howat, 1983; Godsey, 1987; Diwekar and Rubin, 1991). However, there are significant limitations:

1) Monte Carlo integration is inherently approximate, so the reliability estimate is uncertain. However, accurate confidence intervals can be obtained .

2) It can be very time-consuming because a large number of process simulations is required for adequate precision. Confidence intervals decrease very slowly—quadrupling the number of points only halves the confidence interval.

3) Optimization of the control variables z requires an extra iterative level of process simulation, which is time-consuming and might be difficult to perform.

4) For any change in the parameter uncertainties (e.g., sensitivity analyses), the entire integration must be repeated.

Table II.D.2.a-1
Monte Carlo Integration Algorithm to Estimate Design Reliability (Probability Statistics)

1) Set z to its known or best-estimate optimum .
2) Generate random number(s) θ^{Sim}, according to the probability distribution for θ.
3) Perform a process simulation, i.e., solve for x where: $\mathbf{h}\left(d, z, x, \theta^{Sim}\right) = 0$
4) Calculate the active inequality constraint value: $g' = \underset{i=1\ldots m}{Max}\ g_i\left(d, z, x, \theta^{Sim}\right)$
5) If the inequality constraints are not violated $\left(g' \le 0\right)$, θ^{Sim} is inside the boundary so there is no need to optimize z. Skip to (7).
6) Vary z within its operating range ($z \in \mathbf{T}$), and repeat steps 3 and 4 until a case is found where $g' \le 0$ (in which case, θ^{Sim} is inside the boundary region), or it is determined that $g' > 0$ for all values of z (θ^{Sim} is outside the boundary region).
7) Keep track of the number of points inside the boundary region. Calculate $\Re$ and its confidence interval (calculations developed in Appendix B). Go to (1) and continue loop until the confidence interval is sufficiently small.

II.D.2.b. Kubic and Stein Method

First, the probability of failure of each individual inequality constraint is estimated by quadratic approximation and extrapolation. Then, the overall reliability is bounded by use of statistical axioms. This is far quicker than Conventional Monte Carlo. However, the results are imprecise and sometimes inaccurate.

Kubic and Stein (1988) proposed a simplification from statistical theory: Calculate the probability of violating each inequality constraint separately. From this, place min/max bounds on the overall probability using fundamental statistical axioms as follows:

$$\cup_j \ = \ \Pr\left\{ g_j > 0 \right\} \qquad \text{Eq II.D.2.b-1a}$$

$$\sum_{j=1}^{m} \cup_j \ \ge \ \cup \ \ge \ \underset{j=1\ldots m}{Max}\left(\cup_j\right) \qquad \text{Eq II.D.2.b-1b}$$

where

- g_j is the inequality constraint j
- $\cup_j$ is the probability of violating inequality constraint j
- $\cup$ is the overall probability of violating any constraint, i.e., the design unreliability
- m is the number of inequality constraints

The design reliability and unreliability are related by:

$$\mathfrak{R} \;=\; 1 - \bigcup \qquad \text{Eq II.D.2.b-2}$$

or:

$$1 - \sum_{j=1}^{m} \bigcup_j \;\le\; \mathfrak{R} \;\le\; 1 - \underset{j=1\ldots m}{Max}\left(\bigcup_j\right) \qquad \text{Eq II.D.2.b-3}$$

Calculating the probability of failure of each of the individual constraints requires accurate boundary information. But since the final result is approximate, some simplifications are acceptable. By finite difference simulation at $\boldsymbol{\theta}^{Des}$, Kubic and Stein (1988) fitted the inequality constraint to a quadratic function or reduced-order model (Kubic, 1989), then extrapolated to the boundary. This method is simple, but inherently approximate with unknown accuracy.

A major problem with this approach is that even if the approximations could be eliminated, the min/max bounds might be too wide to be useful.

II.D.2.c. Alternative Methods for Linear Process Models

Two alternative methods have been developed. However, they are mathematically complex and unacceptably limited to linear models.

Some researchers have recently attempted to extend the work of Grossmann and coworkers (section II.D.1.) to probability statistics. Pistikopoulos and Mazzuchi (1990) gave credit to Kubic and Stein (1988), but incorrectly noted that their method cannot handle control variables. Straub and Grossmann (1990) mention Kubic and Stein (1988) only in passing. Both research groups performed complex mathematical feats to overcome the difficulties, but the net result is severely limited:

(1) Gaussian quadrature is the integration technique, which is computationally expensive and limited to a maximum parameter dimension of about six (Deák, 1988).

(2) The Pistikopoulos and Mazzuchi (1990) procedure is limited to normal probability distributions.

(3) Process models must be linear.

Fundamentally, these approaches are doomed by the mathematical difficulties and limitations.

II.D.3. Possibility Statistics

The only available calculation procedure is that of Kubic and Stein (1988). It requires nonlinear optimization. Since this case is not studied in this research, no technical development is attempted. This section merely presents a visual picture to qualitatively describe the problem.

Possibility distributions can be viewed similarly to probability distributions. Assuming for convenience that all possibility distributions are triangular (see Figure II.C.3-1, right figure), then the Gaussian distribution in Figure II.D.2-1 is replaced by a multidimensional pyramid, see Figure II.D.3-1. Of course, any distribution may be used instead. The possibilistic parameters are denoted $\boldsymbol{\theta}^{F}$.

The pyramid distribution in Figure II.D.3-1 can be interpreted as follows: The actual parameter value could be anywhere within the base, but it is more 'possibly' in the regions of greater height. The possibility that any value is correct is equal to the height of the distribution at that point.

For the distribution shown in Figure II.D.3-1, there is no overlap with the constraint boundary, so there are no values for which the design can fail. Therefore, this particular design is 100% reliable.

If the possibility distribution overlaps with the constraint boundary, as shown in Figure II.D.3-2, then the maximum height of the distribution outside the boundary is the possibility of failure—point a. And the maximum height of the distribution inside the boundary is the possibility of success—point b. Note that the design reliability and unreliability do not, in general, sum to one.

Calculation of the design reliability requires a nonlinear optimization search to find the boundary point of maximum possibility. Kubic and Stein developed a fairly efficient procedure for this, however it requires several process simulations.

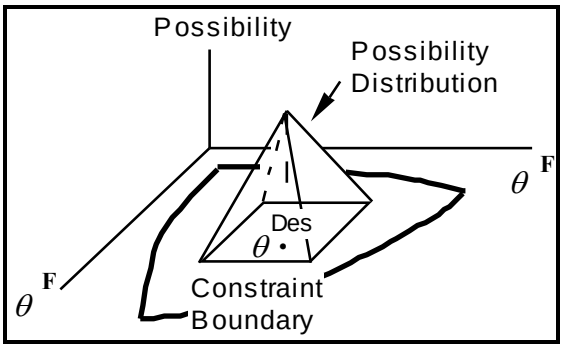


Figure II.D.3-1
Possibility Distribution in the Design Region - 100% Reliable Case

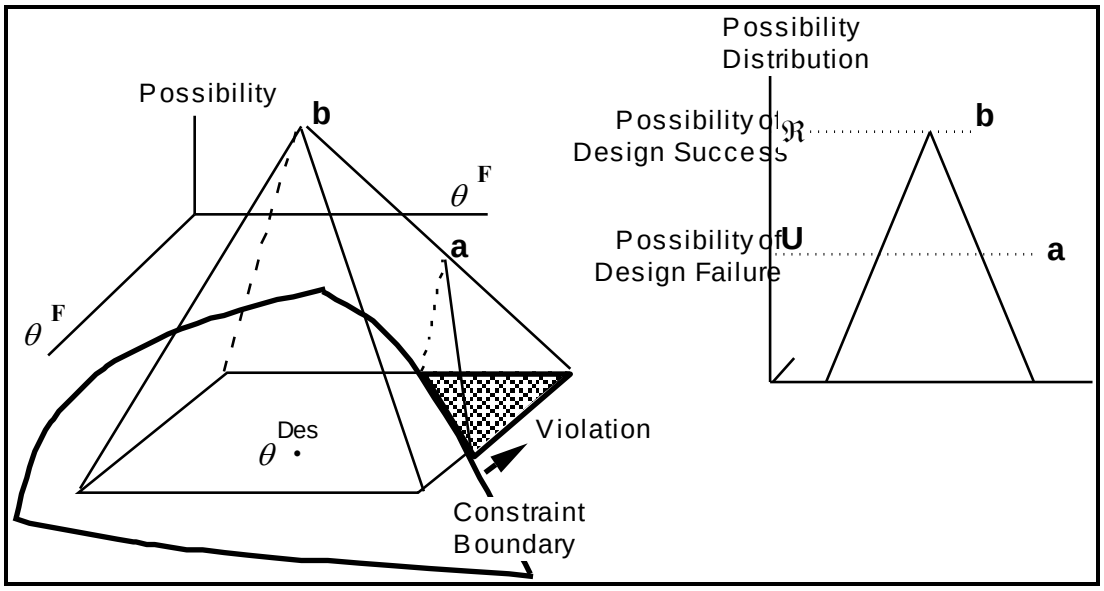


Figure II.D.3-2
Possibility Distribution in the Design Region - Less than 100% Reliable Case

II.D.4. Hybrid Statistics

The only available calculation procedure is that of Kubic and Stein (1988). However, it is imprecise and extremely expensive to perform. This section develops the visual picture without technical development.

The most general situation is when some parameters have probability distributions and some have possibility distributions, a “hybrid” distribution (Kaufmann and Gupta, 1985). It can be visualized by superimposing the two distributions, as shown in Figure II.D.4-1. The possibility distribution (pyramid) has a fixed shape that ‘floats’ on the probability distribution (Gaussian ‘dome’), with position determined by the probability distribution.

The design reliability is the ‘average’ of the possibility distribution, as weighted by the probability distribution. That is, the reliability and unreliability for the possibility distribution (the overlap of the possibility distribution with the constraint boundaries, as discussed in the previous section), are averaged by the probability of that particular possibility distribution occurring.

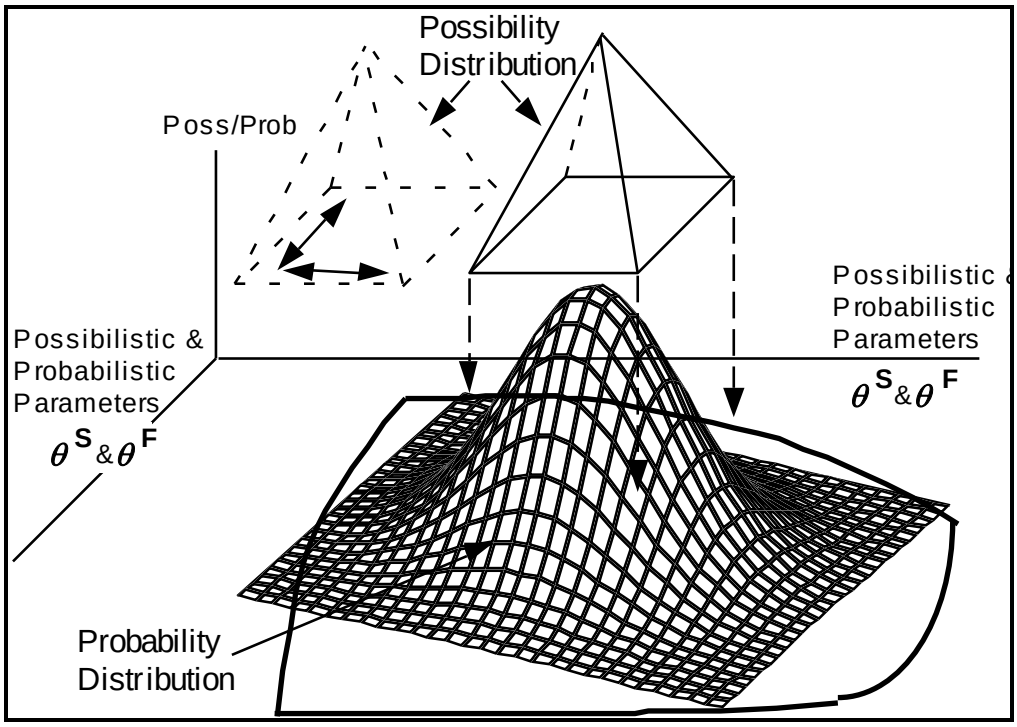


Figure II.D.4-1
Hybrid Distribution in the Design Region

To calculate the design reliability, Kubic and Stein (1988) resort to Monte Carlo simulation. First, the probabilistic parameters are randomly simulated. Then, the design reliability is calculated for the remaining possibilistic case as described in the previous section. This is repeated 50 or 100 times and the results averaged. Obviously, this approach is extremely expensive. Also, even 100 Monte Carlo repetitions is still quite imprecise (>5% uncertainty), which Kubic and Stein do not take into account.

II.D.5. Limitations of the Current Methods

Previous research has been limited to calculating reliabilities when the uncertainties are described by min/max bounds or probability distributions. Not only is this statistically inadequate for chemical engineering problems, but the computations are difficult and/or time-consuming. Hybrid statistics are the state of the art, however Kubic and Stein's (1988) computational approach is prohibitively expensive. The goal now is to make these calculations accurate, precise and efficient.

III. Calculation of Design Reliability by Boundary-Approximation Technique

This section develops my procedure to calculate design with reliability. First, the overall algorithm is presented in section III.A. While the concept is simple, the mathematical details are tricky and are developed in the following sections.
The procedure develops a mathematical approximation of the constraint boundary by finding points on the boundary and connecting them using a form of linear interpolation. The calculation of design reliability and other statistical analyses then becomes straightforward. The major computational effort is spent finding the boundary points. The procedure is iterative to minimize computational effort.

III.A. Overall Procedure

As described in the previous section, the different statistical distributions (probability, possibility, or hybrid) require different computational procedures. However, in all cases, the computational bottleneck is that the process simulations are *inside* the statistical loop. The only way to break this bottleneck is to decouple the process simulations from the statistics. This can be accomplished by recognizing the common requirement of all statistical techniques—knowing the whereabouts of the constraint boundary. If the constraint boundary is accurately known, the statistical computations become trivial and very fast. This suggests a two-part algorithm: (1) Map the constraint boundary. (2) *Then*, perform the statistical computations. Computational effort inside the statistical loop is replaced by computational effort to map the constraint boundary.

Table III.A-1
Comparison of Existing and New Algorithms to Calculate Reliability

Existing Approach	New Approach
1) Statistical Calculation Loop a) Do Process Simulation End Loop	1) Map the Constraint Boundary 2) Perform the Statistical Computations

The constraint boundary can be mapped by finding several points on the boundary then interpolating among them. However, interpolation is inherently inaccurate, so some error is inevitable. Of course, the more boundary interpolation points there are, the better the boundary-approximation. On the other hand, to minimize the computational effort it is desirable to find as few boundary points as necessary.

This conflict may be resolved by using an iterative approach: (1) Start with a few 'key' boundary points to calculate an approximate design reliability. (2) Then, find additional boundary points in the most sensitive and uncertain locations and recalculate the reliability. (3) Continue until the reliability estimate is sufficiently accurate. The algorithm is summarized in Figure III.A-1.

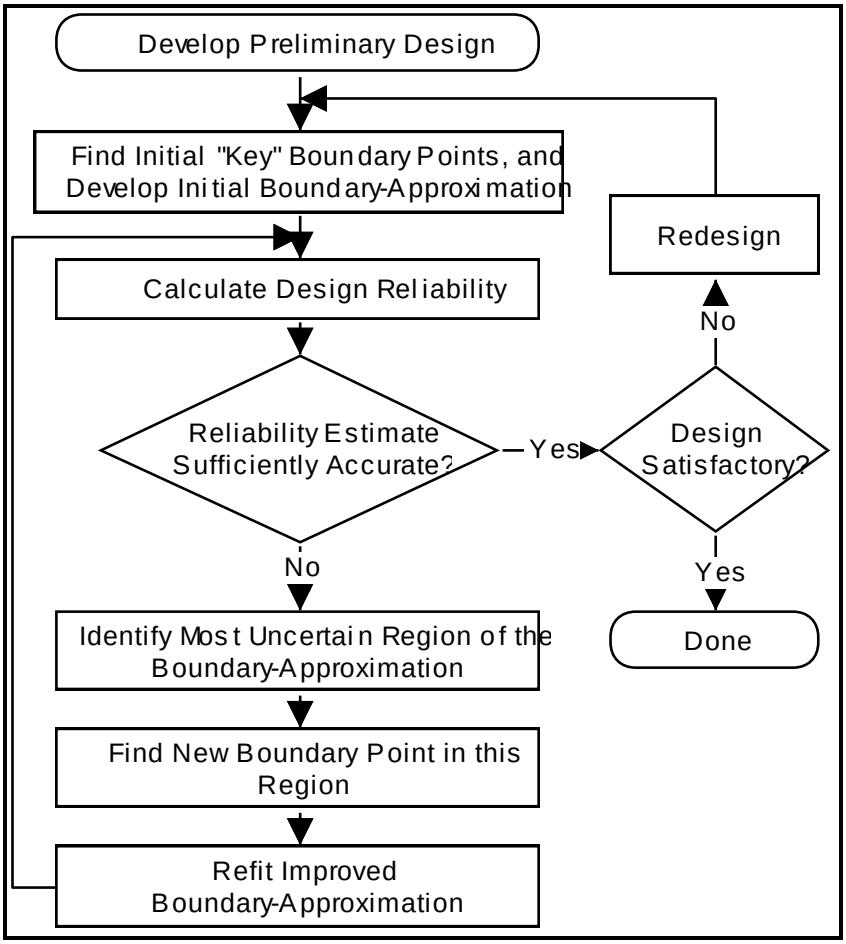


Figure III.A-1
Flowchart of Design Evaluation and Reliability Calculation

The designer is an integral part of this algorithm. First, he or she develops the preliminary design. Next, the iterative procedure must be monitored to determine when the reliability estimate is sufficiently precise to accept or reject the design. Poor designs will be identified very quickly and do not require precise reliability estimates. The designer must then alter the design and try again. Finally, the designer obtains a satisfactory design and allows the reliability calculation procedure to converge with sufficient accuracy and precision.

Section III.B describes the mathematics of constructing the boundary-approximation from the individual boundary points. Section III.C describes the calculation of the reliability estimate. Section III.D describes how to identify the most uncertain boundary region for further investigation. Finally, section III.E describes two different algorithms for finding additional boundary points.

III.B. Boundary-Approximation Mathematical Techniques

This section details the mathematical development of the boundary-approximation. Discussion is organized into two parts: (1) Subdivision of the parameter-space into the set of 'gaps' or regions between boundary points, each of which may then be treated separately; (2) Optimistic, pessimistic, and best-estimate boundary-approximations for each region.

The boundary-approximation is handled by subdividing the parameter-space into the set of 'gaps' or empty regions between the boundary points, then analyzing each region individually. First, the *axial* boundary points are found, i.e., in the positive and negative direction of each parameter dimension, to ensure that the entire parameter-space is covered to begin with. Thereafter, additional boundary points are handled by finding which region they fall into, then subdividing that region into smaller pieces. This can continue indefinitely.

Local curvature is determined from the point+gradient information of each region's boundary points. Linear interpolation plus curvature information provides best-estimate, optimistic and pessimistic estimates for each region's boundary.

Parameter-space subdivision is treated in section III.B.1. Local curvature boundary analysis is treated in section III.B.2.

III.B.1. Subdivision of the Parameter-Space into Local Boundary Regions

III.B.1.a. Initialization Using the Axial Boundary Points

To ensure that the boundary-approximation is completely covered to begin with, it is initialized with the *axial* points, i.e., the set of boundary points in both directions along each parameter axis with the focalpoint or origin at θ^{Des}. This is shown schematically in Figure III.B.1.a-1. For p parameters, there are 2p axial boundary points. The algorithm for finding these points is developed later in section III.E.

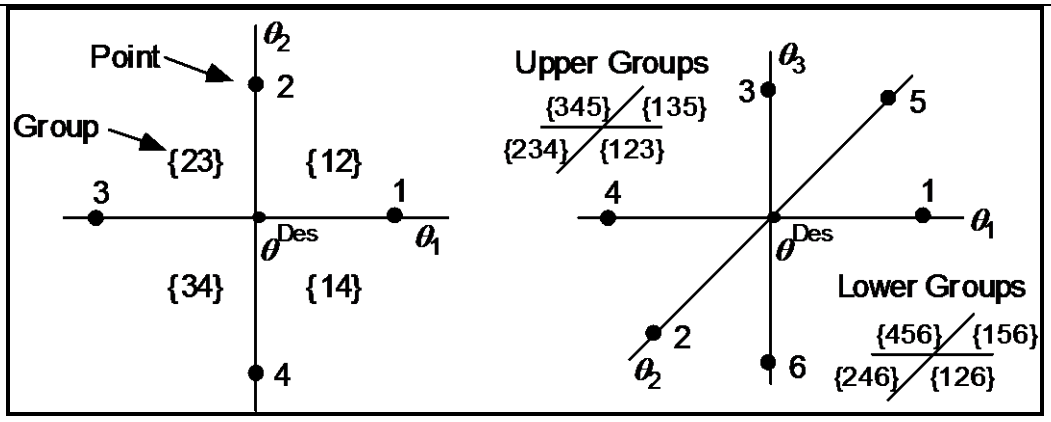


Figure III.B.1.a-1
Axial Boundary Points in 2-D and 3-D

The 'gaps' between points are called *nearest-neighbor groups*. Initially, this is simply the 2^p set of hyperquadrants. Figure III.B.1.a-1 shows these groups within parentheses {…}. Each nearest-neighbor group is composed of p boundary points.

Watson (1988) provides the following equation to create the initial set of nearest-neighbor groups. The axial points are first sorted by: all positive directions ($1 \ldots p$), then all negative directions ($1 \ldots p$). They are numbered $1 \ldots 2p$. Then, Eq III.B.1.a-1 provides the correct neighbor groupings:

For group $i = 1 \ldots 2^p$; For point $j = 1 \ldots p$;

$$\text{(group } i\text{, point } j\text{)} = j + \frac{p}{2} - \frac{p}{2} \times (-1)^{\left[int\left(\frac{i-1}{2^{(p-j)}} \right) \right]} \qquad \text{Eq III.B.1.a-1}$$

where $int()$ is the function: greatest integer not greater than the argument

III.B.1.b. Adding New Boundary Points

Since the parameter-space is completely covered after initialization, any additional boundary point must fall inside one of the existing nearest-neighbor groups, thus subdividing the region still further. When a new boundary point is added, it destroys the old neighbor group, and creates p new ones. The new groups consist of the new point combined with the p different sets of p-1 points from the old group. For example, say a new point (#7) falls inside the old neighbor group {123}, as shown in Figure III.B.1.b-1. Then, {123} is destroyed to create new groups {127}, {237}, {137}.

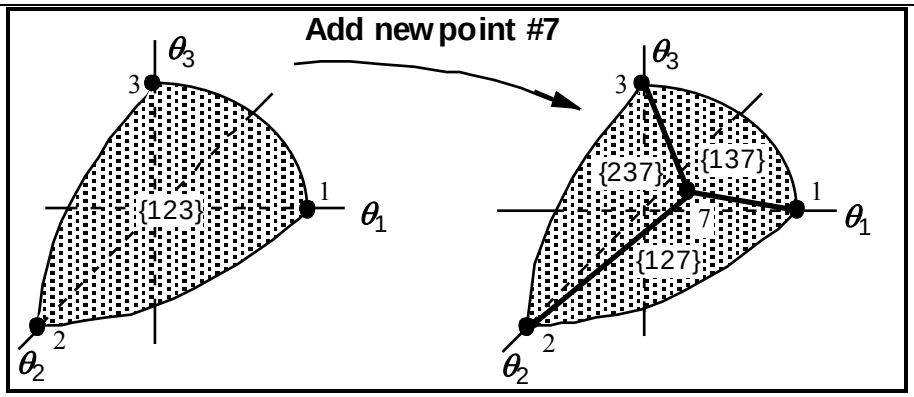


Figure III.B.1.b-1
Addition of New Boundary Point Destroys Old Nearest-Neighbor Group and Creates P New Ones

The procedure must be able to determine in which nearest-neighbor group the new point lies. Unfortunately, Watson's (1988) method was unacceptable because it allows overlap

between groups. So a new method was developed by defining exact borders between the neighbor groups, see Figure III.B.1.b-2. Each group has p sides or *borders*. Each border is a hyperplane that connects p-1 of the group's points and the focalpoint $\boldsymbol{\theta}^{Des}$. A new point is inside a neighbor group if it is inside every one of the group's p borders.

Each border equation may be solved using the form:

$$\mathbf{A} \cdot \mathbf{a} + \mathbf{1} = \mathbf{0} \qquad \text{Eq III.B.1.b-1}$$

where:

$\mathbf{A}$ is a [p·p] array, one row being the coordinate values of $\boldsymbol{\theta}^{Des}$, and the other p-1 rows being the coordinate values of the connecting boundary points

$\mathbf{a}$ is the [p·1] vector of constants being solved for that make up the equation of the border hyperplane

$\mathbf{1}$ is a [p·1] vector of 1's

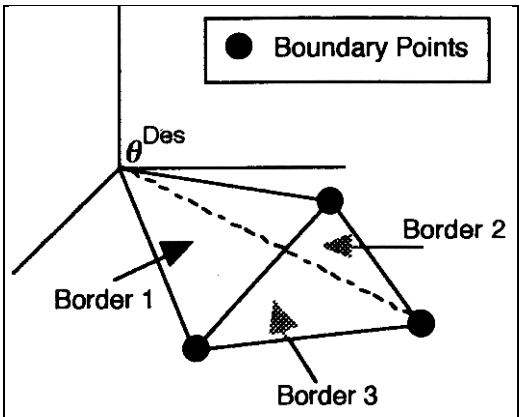


Figure III.B.1.b-2
Neighbor Group Boundaries Made Up of P Bordering Planes

Thus, each border requires solution of p linear equations in p unknowns. A simple matrix inversion supplies the solution:

$$\mathbf{a} = -\mathbf{A}^{-1} \cdot \mathbf{1} \qquad \text{Eq III.B.1.b-2}$$

This calculation can fail on certain rare occasions due to matrix singularity, for example when one of the boundary points is at the origin (0,0,…,0), or when points are collinear. However, these unusual situations will not normally occur.

Note that the borders for the axial point initialization set are simply the set of axial planes, which do not need to be calculated by matrix inversion. This is important because the initial computational effort would be significant otherwise. For instance, if there are 10 parameters, there are 2^{10} = 1024 nearest-neighbor groups initially, which would require 10240 inversions of a [10·10] matrix. On the other hand, additional boundary points require little computational effort because each new point requires only p^2 matrix inversions of order p.

To determine if any point ($\boldsymbol{\theta}^{Sim}$) is inside or outside the border, perform the calculation:

$$\mathbf{a}^T \cdot \boldsymbol{\theta}^{Sim} + 1 \quad = \quad \sum_{i=1}^{p} a_i \cdot \theta_i^{Sim} + 1 \quad = \quad ? \qquad \text{Eq III.B.1.b-3}$$

The sign of the remainder indicates which side of the border the point is on (a result of zero indicates the point is exactly on the border). Unfortunately, which side is which is unknown a priori. However, this is remedied by first performing the calculation using the remaining boundary point of that particular group (the one not connected to the current border) to determine the sign, since we know that point is definitely inside the border. For example, in Figure III.B.1.b-1 left-hand side, point 1 is inside the 2-3 border of group {123}.

Arbitrarily, the remainder of Eq III.B.1.b-3 is negative when the point is inside the border. Therefore, if the check point gives a positive result, the equation is reversed, i.e.:

$$\text{Sign Reversal:} \quad \mathbf{a}^T \cdot \boldsymbol{\theta}^{\text{Sim}} + 1 \;\Rightarrow\; -\mathbf{a}^T \cdot \boldsymbol{\theta}^{\text{Sim}} - 1 \qquad \text{Eq III.B.1.b-4}$$

Note that the border matrix inversion (Eq III.B.1.b-2) is performed only once when the new neighbor group is created. Thereafter, to determine if a point is inside the neighbor group, all that is required is a multiplication + summation p-loop, done on average $p/2$ times per neighbor group.

While this calculation is very quick, it must be done a large number of times as dimension increases—about $2^p p/4$ border calculations per additional point—which creates a computational bottleneck. This can be avoided by a shortcut: First, organize the groups according to the hyperquadrant they reside in. Then, for each new point, identify its hyperquadrant first, then check only the neighbor groups in the same hyperquadrant. This is possible because there is no overlap, i.e., no group lies in more than one hyperquadrant. At initialization, the groups are uniquely numbered from 1 to 2^p. Thereafter, each new group takes on the same number as the old group it replaces. Using this approach, the number of border calculations per additional point is on the order of $(\#\text{ groups in hyperquadrant}) \times p/4$.

III.B.1.c. Disadvantages of Approach

Subdividing the parameter-space into the set of nearest-neighbor groups allows the boundary surface to be analyzed in small portions. While this was necessary to make the interpolation problem manageable, it is not the most powerful approximation method because it does not always provide the *nearest* boundary points from which to interpolate. Consider a 3-D example in Figure III.B.1.c-1 where four boundary points (1…4) are connected into two nearest-neighbor groups ({123}, {134}). The addition of a new point (5) results in the grouping shown at the top. However, a potential group {345} was missed, as shown at the bottom. If {345} is included, note that it overlaps two other groups {135} and {145}, which is undesirable. However, group {345} is probably a better interpolator for that region.

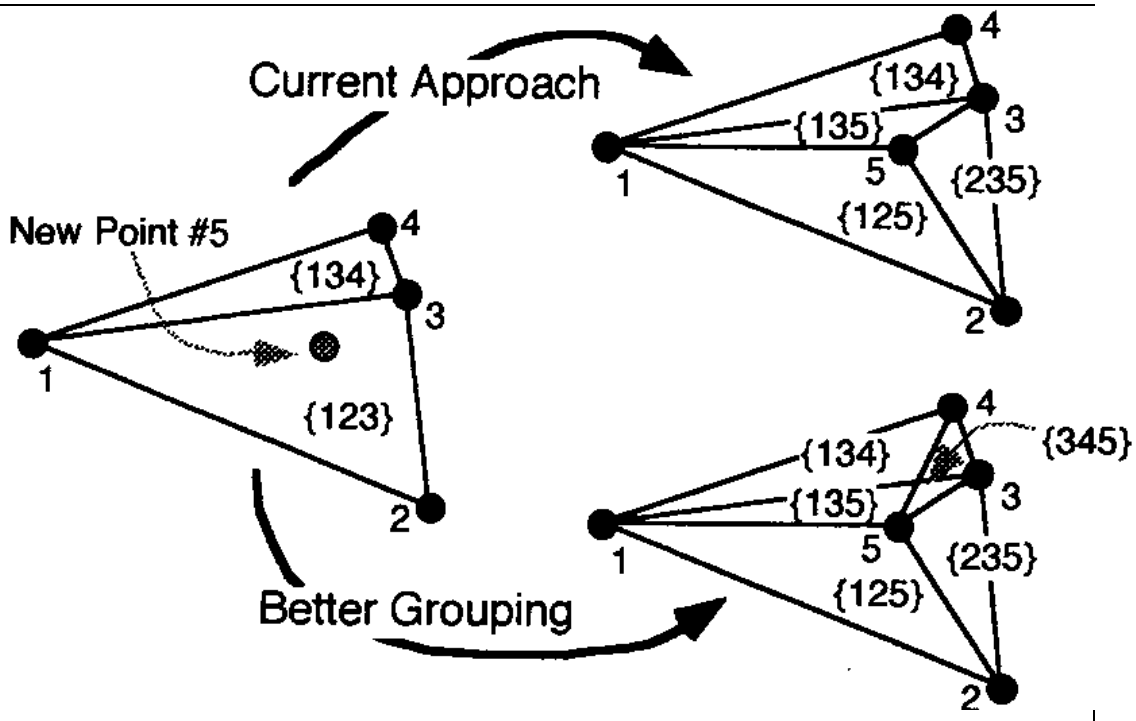


Figure III.B.1.c-1
Current Parameter-Space Subdivision Algorithm Does not Yield Optimum Nearest-Neighbor Grouping

For this research, the loss in interpolation power was considered acceptable to ensure nonoverlap between nearest-neighbor groups. Better surface approximation methods can be developed in the future.

A second disadvantage is that the procedure cannot handle additional boundary points that lie *exactly* on a neighbor group's border. This occurs whenever an additional boundary point has one or more individual parameter value(s) at the design or focalpoint value(s). While the procedure will not generate these points during convergence, these points may often be very useful for improving accuracy at initialization, e.g., see section V.B.3.a. This is because border points may be used to improve more than one nearest-neighbor groups at once.

Figure III.B.1.c-2 shows an example where four points (1…4) connect to form two neighbor groups {123} and {124}. An additional point (5) falls exactly on the border between points 1 and 2. Proper treatment would include this point in both existing neighbor groups. However, the current algorithm has no provision for this. For this research, a quick workaround was used—simply duplicate this point with small offsets to place separate copies of the point into each neighbor group, as shown in the Figure (points 5a and 5b). Better

treatment must come later, as this approach becomes very cumbersome and causes computer roundoff as parameter dimension increases.

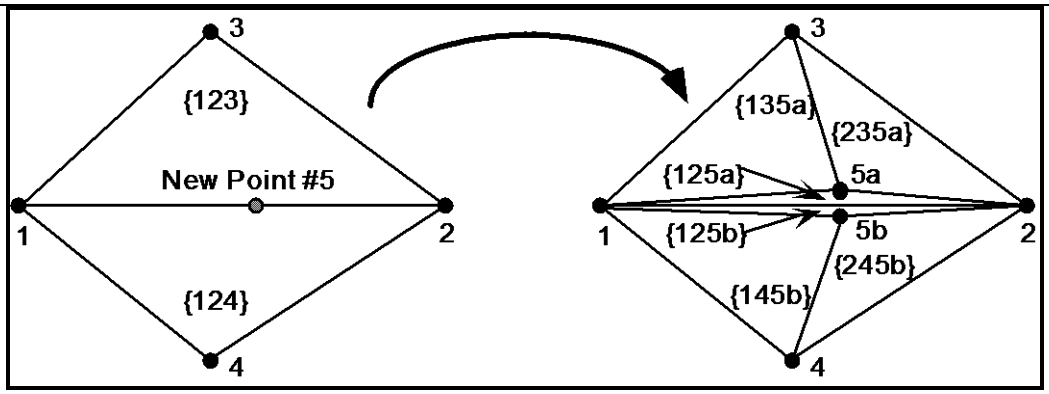


Figure III.B.1.c-2
Simple Algorithm Workaround: Additional Boundary Points That Lie Exactly On The Neighbor Group Border-Planes Are Duplicated With Small Offsets So As To Fall Into Each Relevant Group

III.B.2. Analysis of Local Boundary Regions

After the parameter-space is subdivided into the set of nearest-neighbor groups, each group is analyzed separately to develop optimistic, pessimistic, and best-estimate boundary-approximations for that region. First, two geometric boundary-approximations are developed: (1) Tangent-planes boundary, and (2) Connecting-plane boundary. The tangent-planes boundary is the intersection of the tangent-planes at each of that group's boundary points. (Note: The type of intersection depends on curvature, as discussed later.) The connecting-plane boundary simply connects the group's boundary points by linear interpolation. Next, the point+gradient information at each boundary point of the neighbor group is analyzed to determine local boundary curvature—convex, concave, or saddle. Curvature determines which of the tangent-planes and connecting-plane boundaries is optimistic and which is pessimistic. The best-estimate boundary-approximation is always the tangent-planes boundary.

A feature not yet introduced is that the boundary-approximation procedure uses two types of boundary points: (1) Failure points, and (2) Nonfailure points. A failure point lies exactly on the constraint boundary. However, a nonfailure point is inside the constraint boundary. The latter occurs when the boundary is so far away that it is difficult and wasteful to search for it. Then, it is more efficient to truncate the search at some maximum statistical distance. The boundary-approximations are constructed by including the nonfailure point in the connecting-plane boundary, but not including it in the tangent-planes boundary (because it has no tangent).

III.B.2.a. Creating Connecting-Plane and Tangent-Plane Boundary-Approximations

Figure III.B.2.a-1 shows a schematic of the connecting-plane and tangent-planes boundary-approximations in 2-D. A connecting plane boundary is created simply by connecting all boundary points with straight lines (hyperplanes). A tangent-plane boundary is the intersection of the tangent-planes at each boundary point.

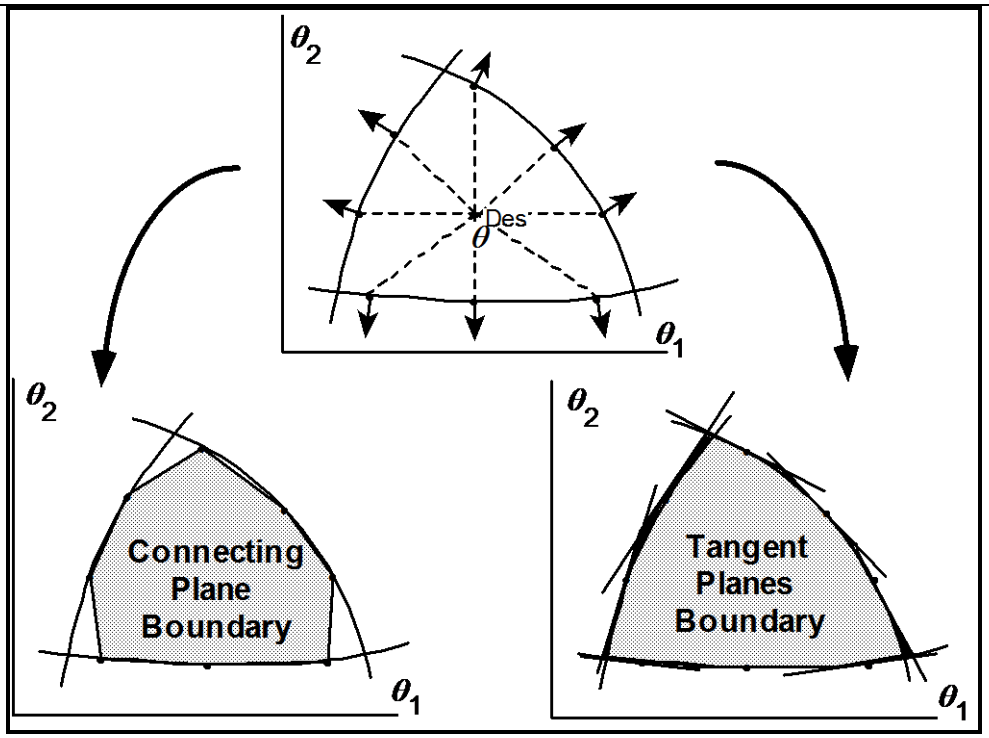


Figure III.B.2.a-1
Schematic of Connecting-Plane and Tangent-Plane Boundary-Approximations

Connecting-Plane Mathematics

For each neighbor group, the connecting-plane boundary is the hyperplane connecting the p points of that group, see Figure III.B.2.a-2. Since there are p points in p-space, the solution is unique and can be solved by p equations in p unknowns:

$$\mathbf{B}\cdot\mathbf{b}+\mathbf{1}=0 \qquad \text{Eq III.B.2.a-1}$$

where

$\mathbf{B}$ is a [p·p] array, the i'th row being the coordinates of the i'th point in the neighbor group (i=1…p)

$\mathbf{b}$ is the [p·1] vector of constants being solved for, i.e., that make up the connecting-plane equation

1 is a [p·1] vector of 1's

The constants in $\mathbf{b}$ are solved by inverting $\mathbf{B}$:

$$\mathbf{b} = -\mathbf{B}^{-1} \cdot \mathbf{1} \qquad \text{Eq III.B.2.a-2}$$

The matrix inversion fails if any of the points are collinear, if two or more points are duplicated, or if a point is on the origin. This is unlikely in practice.

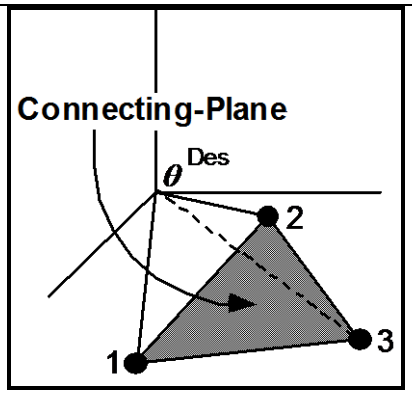


Figure III.B.2.a-2
Connecting-plane in 3-D

To determine if a point (θ^{Sim}) is inside or outside the connecting-plane, we calculate $\mathbf{b}^T \cdot \theta^{Sim} + \mathbf{1} = ?$ to determine the sign. However, as for the border-planes (section III.B.1.b), determining which side is which requires a check point, for which θ^{Des} is suitable because it is always inside the connecting-plane. Arbitrarily, the sign is negative when a point is inside the connecting-plane boundary.

Tangent-Plane Mathematics

The equation for each boundary point's tangent-plane is as follows, see Figure III.B.2.a-3:

$$\left(\nabla \hat{\theta}^i\right)^T \cdot (\theta - \hat{\theta}^i) = 0 \qquad \text{Eq III.B.2.a-3}$$

where

$\hat{\theta}^i$ is the coordinates of boundary point i, dimension [p·1]

$\nabla \hat{\theta}^i$ is the gradient at boundary point i, dimension [p·1], where the j'th vector component is $\frac{\partial g}{\partial \theta_j}$

θ is any point on the tangent plane

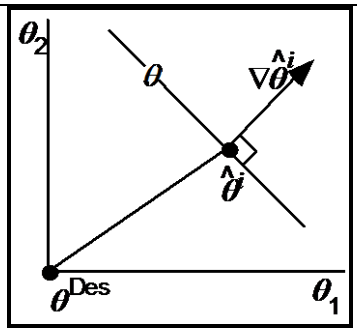


Figure III.B.2.a-3
Tangent Plane
Boundary Schematic

To determine if a point θ^{Sim} is inside or outside this boundary, θ^{Sim} is simply inserted into Eq III.B.2.a-2 in place of θ. If the result is negative, then θ^{Sim} is inside the tangent-plane boundary. Note: The result is negative if the gradient vector points *away* from the focalpoint (θ^{Des}), which should always be the case. But to be certain, the program checks this by performing the calculation with θ^{Des}.

III.B.2.b. Boundary Curvature Analysis

Next, each neighbor group is analyzed to determine the local curvature, which may be convex, concave, or saddle. This is determined by checking each point in the group to see if it is: (1) inside the tangent-planes of all other points in the group, which indicates convexity, (2) outside the tangent-planes of all other points in the group, which indicates concavity, (3) inside some, but outside others, which indicates more complex saddle shapes. The algorithm is shown in Figure III.B.2.b-1.

Note that because of the limited information, concavity or convexity cannot be guaranteed. However, this becomes more likely as the neighbor groups become smaller.

When neighbor groups contain a nonfailure boundary point(s), the tangent-plane analysis for this point is skipped (because there isn't one). However, it is determined if the nonfailure point is inside or outside the tangent-planes of the other failure points.

Another useful check is to determine whether or not the neighbor group has the same active constraint or different ones. If the former, then we can have greater confidence that the curvature information is valid. If the latter, then the boundary approximation is more uncertain since it has first-order discontinuity.

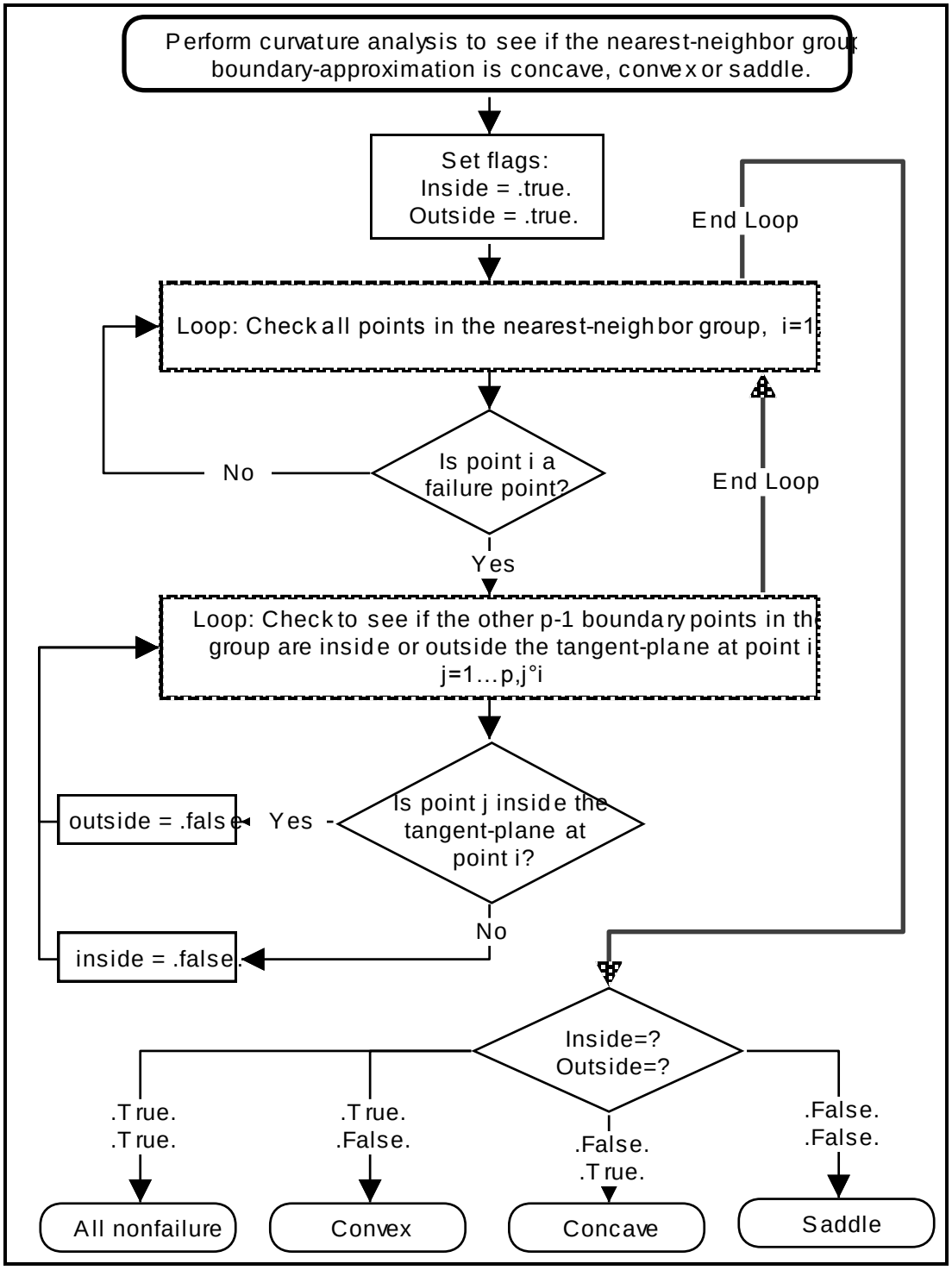


Figure III.B.2.b-1
Boundary Curvature Analysis Algorithm for Each Nearest-Neighbor Group

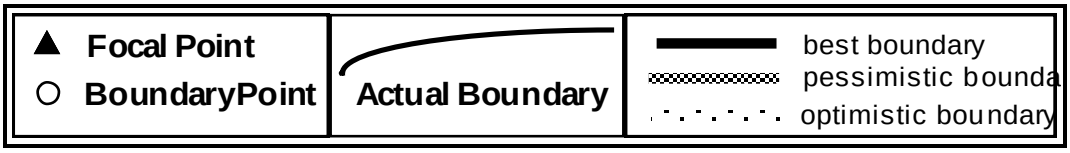
III.B.2.c. Best-Estimate, Optimistic and Pessimistic Boundary-Approximations

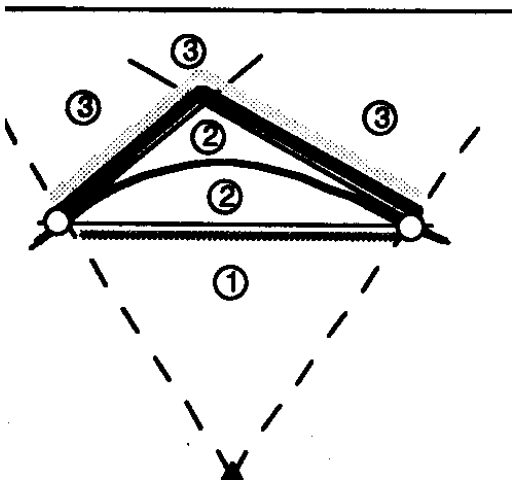
A tangent-plane boundary is always more accurate than the connecting-plane boundary. This is because, while both are accurate at the boundary points, from geometry the tangent-plane is asymptotically accurate near the boundary points, while the connecting-plane is immediately inaccurate away from the boundary points. Therefore, regardless of the boundary curvature, the tangent-planes boundary is always a better choice as the *best-estimate* boundary-approximation.

Depending on curvature, different boundary-approximations are optimistic and/or pessimistic. Also, the tangent-plane boundary uses different intersections. Each case is now discussed.

Convex Curvature

Figure III.B.2.c-1 shows a 2-D schematic. Here, the connecting plane is pessimistic and the tangent-planes boundary is optimistic. The tangent-planes boundary (and the best-estimate boundary) is the internal intersection of the individual point tangent-planes, i.e., a point is inside the tangent-planes boundary if it is inside *all* of the tangent-planes of that group's points.





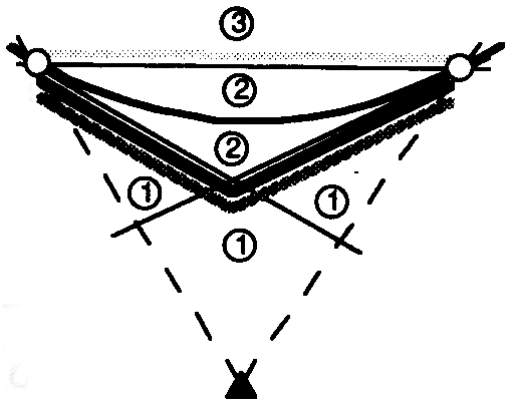
Loc.	Best	Opt.	Pess.
1	In	In	In
2	In	In	Out
3	Out	Out	Out

Inside Tangent-Planes Boundary [=]
Inside ALL Individual Neighbor
Point's Tangent-Plane

Figure III.B.2.c-1
Optimistic, Pessimistic, and Best-Estimate Boundary-Approximations for Convex Curvature

Concave Curvature

Here, the situation is reversed, see Figure III.B.2.c-2. The tangent-planes boundary is pessimistic, and the connecting plane is optimistic.41The tangent-planes (and best-estimate) boundary is the internal union of the individual point tangent-planes, i.e., a point is inside the tangent-planes boundary if it is inside *any* individual point tangent-plane.



Loc.	Best	Opt.	Pess.
1	In	In	In
2	Out	In	Out
3	Out	Out	Out

Inside Tangent-Planes Boundary [=]
Inside ANY Individual Neighbor
Point's Tangent-Plane

Figure III.B.2.c-2
Optimistic, Pessimistic, and Best-Estimate Boundary-Approximations for Concave Curvature

Saddle Curvature

Saddle curvature is a problem because the uncertain curvature makes it impossible to define rigorously optimistic and pessimistic boundaries, and the tangent-planes boundary is systematically in error, as shown in Figure III.B.2.c-3. Both individual tangent-planes are in error at the other boundary point and no intersection or union can correct for this.

There is no ideal solution using only linear boundaries. However, a reasonable approach is to define the best-estimate boundary as the tangent-plane of the *nearest* boundary point. This eliminates the systematic error at the boundary points, but at the cost of step changes in the middle, as shown in Figure III.B.2.c-4. While the step change is certainly inaccurate, at least it occurs in the region of greatest uncertainty (farthest from any boundary point).

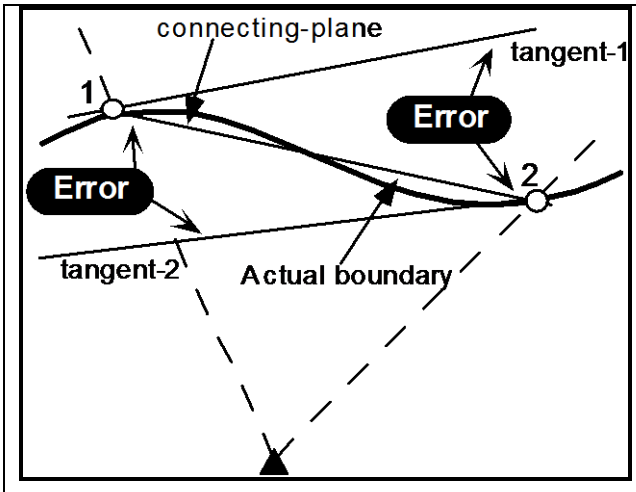
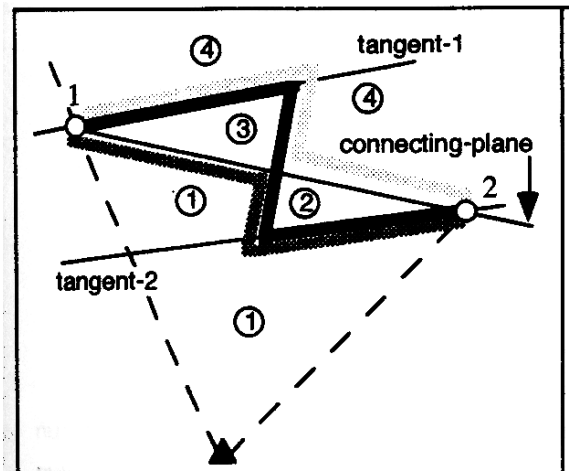


Figure III.B.2.c-3
Saddle Curvature - *Incorrect* Approach



Loc.	Best	Opt.	Pess.
1	In	In	In
2	Out	In	Out
3	In	In	Out
4	Out	Out	Out

Inside Tangent-Planes Boundary [=] Inside The Nearest Individual Neighbor Point's Tangent-Plane

Figure III.B.2.c-4
Optimistic, Pessimistic, and Best-Estimate Boundary-Approximations for Saddle Curvature

Nonfailure Boundary Point Handling

When a neighbor group contains a nonfailure boundary point(s), the connecting-plane connects to it, and the tangent-planes boundary ignores it. Thus, a nonfailure point adds nothing to the best-estimate boundary-approximation. However, it does have the secondary effect of tightening the optimistic or pessimistic boundary-approximations.

For the special case where the neighbor group consists solely of nonfailure boundary points, there is no tangent-planes boundary and no curvature information, only a connecting-plane. The optimistic and best-estimate boundary-approximations are assumed to be 'very far out', i.e., nonexistent or unbounded, and the pessimistic boundary-approximation is the connecting-plane.

III.C. Design Reliability Estimation Using Monte Carlo Integration With the Boundary-Approximation

This section discusses the estimation of design reliability using the mathematical boundary approximation. Monte Carlo integration is used for the probability distribution case. Since the integration is very fast, the imprecision caused by Monte Carlo integration may be reduced to negligible.

The statistical probability distribution must be integrated throughout the mathematical boundary-approximation to calculate the design reliability. This can be done analytically or by Monte Carlo. Monte Carlo integration is imprecise by nature, while analytical integration is exact. However, analytical integration has several disadvantages:

(1) An exact result has no significant advantage because the boundary-approximation itself is uncertain.

(2) Many probability distributions must be integrated numerically, i.e., limited precision.

(3) The geometrical region of integration is complicated.

(4) As dimension increases, analytical integration becomes very slow (Kalos and Whitlock, 1986). Deák (1988) recommends Monte Carlo integration for dimensions greater than 5.

In contrast, Monte Carlo integration is simple to perform and is unaffected by the number of dimensions. The disadvantage of imprecision is eliminated by generating a large number of random points, which is possible because the boundary-approximation calculation is very fast. For more information on the statistics of Monte Carlo integration, see section II.D.2.a and Appendix B. The remainder of this section discusses its application to the boundary-approximation procedure.

First each random point must be examined to find the nearest-neighbor group in which it resides, as described in section III.B.1.b. Then this point is checked to see if it is inside or outside the best-estimate, optimistic, and pessimistic boundary-approximations of that particular neighbor group, as described in section III.B.2.

As the Monte Carlo integration proceeds, it keeps track of the following information: (1) Number of random points in each neighbor group; (2) number of points in each group outside the optimistic boundary; (3) number of points in each group outside the pessimistic boundary; (4) number of points in each group outside the best-estimate boundary. (It is faster to keep track of the points *outside* the boundary because there are fewer of them, provided $\leftarrow > 50\%$.) The design reliability is estimated by summing over all the nearest-neighbor groups:

$$\mathfrak{R}^{Opt} = 1 - \frac{\#\text{pts outside optimistic boundary}}{M} \qquad \text{Eq III.C-1}$$

$$\mathfrak{R}^{Pess} = 1 - \frac{\#\text{pts outside pessimistic boundary}}{M} \qquad \text{Eq III.C-2}$$

$$\mathfrak{R}^{Best} = 1 - \frac{\#\text{pts outside best-estimate boundary}}{M} \qquad \text{Eq III.C-3}$$

where M is the total number of Monte Carlo points

Random Number Generator

1,000 random points are usually sufficient for initial analyses. However, high precision studies may require 10,000 to 100,000 random points. Each point consists of p random numbers. There are two basic options for handling random number generation:

(1) Generate the random number sequence once and save it to a data file(s). Then, read the data from the file(s) as needed.

(2) Generate the random number sequence by computer every time it is needed.

The computer can generate random numbers much faster than it can read them from a data file. However, there are occasions when data file storage may be faster for the following reason:

The first time the Monte Carlo integration is performed, the entire random number sequence is generated, every point is analyzed, and subtotals are kept for each nearest-neighbor group. However, as each new boundary point is added, only the new nearest-neighbor groups need re-integration—the subtotals for all other groups are unchanged. This re-integration is much faster and may be handled in one of two ways:

(1) Regenerate the entire random number sequence, but only check to see if the points are inside the new neighbor groups.

(2) Store the random numbers in a set of data files—one for each nearest-neighbor group. Then, simply open the data file for the neighbor group being destroyed, recheck these points with the new neighbor groups, then create p new data files for the new groups, deleting the old data file.

Since the number of neighbor groups increases rapidly as parameter dimension increases (2^p at initialization, $+p-1$ for every additional point) , the number of data files may exceed typical disk storage capacity. Therefore, regression without storage was used.

Table III.C-1 lists the approximate time required to perform a 10,000 Monte Carlo integration for the distillation test problem on a Macintosh II computer.

Table III.C-1
Execution Time to Perform Monte Carlo Integration for 10,000 Random Numbers Using the Boundary-Approximation Procedure

Dimension	Initialization (Axial Points Only)			At Convergence (300 Pts Max.)		
	# Bndry Pts	Entire Integration	Incremental Integration	# Bndry Pts	Entire Integration	Incremental Integration
2	4	26 secs	19 secs	50	40 secs	19 secs
5	10	65	50–60	300	370	50–60
10	20	190	160–170	300	410	160–170

Note that time required to perform the entire integration and the incremental integration increases with dimension. However, the incremental integration stays constant as the number of boundary points increases.

The random number generator for the Macintosh computer is tested in Appendix B-1.

III.D. Identification of the Most Uncertain Constraint Boundary Location for Further Investigation

This section discusses the identification of the most statistically uncertain region of the boundary-approximation, and the selection of the radial direction in which to search for the next boundary point. The uncertain boundary region is the geometrical difference between the optimistic and pessimistic boundary-approximations. The most statistically uncertain region is the one containing the most Monte Carlo points.

In the previous section, the optimistic and pessimistic boundary-approximations were developed as upper and lower bounds of the actual boundary. Therefore, the difference between the two serves as a useful measure of the boundary uncertainty. This region has the desirable geometry that it is small at the boundary points and is maximized in the space farthest away from any boundary point, see Figure III.D-1.

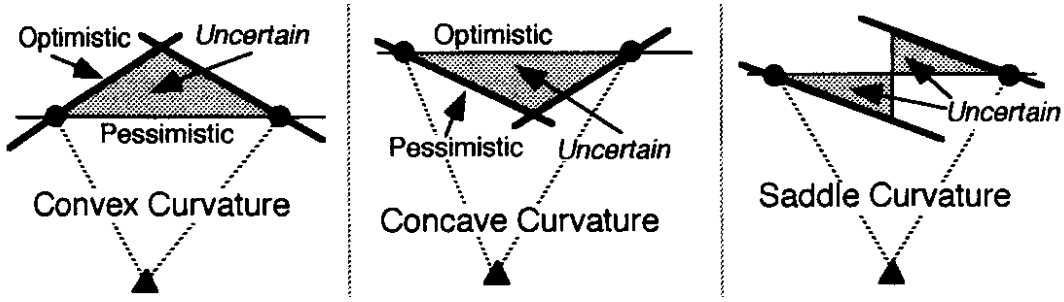


Figure III.D-1
Difference Between the Optimistic and Pessimistic Boundary-Approximations as a Measure of the Boundary Uncertainty

Converging the optimistic and pessimistic boundary-approximations to each other will also converge the best-estimate boundary-approximation. Thus, a useful objective is to minimize the region between the optimistic and pessimistic boundary-approximations—the region of *uncertainty* . However, this should be done statistically and not geometrically, that is, by concentrating only on the most statistically significant uncertain regions.

Each nearest-neighbor group has its own region of uncertainty. The statistical significance of this region is proportional to the number of random Monte Carlo points that fall within it. Therefore, the group with the greatest number of random points in its region of uncertainty is selected as the location for the next boundary point.

A specific radial direction vector must be identified to search for the next boundary point within this neighbor group. Alternatives are:

(1) Search towards the *midpoint* of the nearest-neighbor group.

(2) Search towards the region where the Monte Carlo points are concentrated.

Searching towards the midpoint will cause the greatest geometric improvement. However, it ignores the available statistical information and may have poor convergence. For example, if all the Monte Carlo points are concentrated in one corner of the region of uncertainty, a new boundary point at the midpoint will provide little improvement in the reliability estimate.

Searching in the direction of the greatest concentration of Monte Carlo points can be accomplished by averaging the location of these Monte Carlo points to provide a ‘centered’ location, then searching toward this center.

Both approaches were tested. The second is superior and was used for this research

One significant point should be noted: As dimension (number of parameters) increases, the number of nearest-neighbor groups rises rapidly and causes ‘sparsity’ in the Monte Carlo point mapping. For example, if the Monte Carlo integration is performed for a ten-dimensional case using 10,000 random points, this averages only 10 random points per nearest-neighbor group, with even fewer points in the region of uncertainty. The effort then becomes directed towards resolving individual Monte Carlo points instead of mapping the boundary which may be inefficient. At this point, the number of Monte Carlo points should be increased.

Alternatively, the Monte Carlo points in the region of uncertainty can be resolved individually by Conventional Monte Carlo simulation when the number drops below a certain value. Clearly, if 10 simulations are required to obtain a new boundary point, then if the number of uncertain Monte Carlo points in that group is less than 10, conventional simulation would be more efficient.

III.E. Line-Search Algorithms to Find the Boundary Point in any Specified Radial Direction

Two approaches were developed to search for the constraint boundary in any given radial direction: (1) A nested-loop algorithm, which requires no process simulator modification but is slow, and (2) a simultaneous algorithm, which is faster but requires simulator modification.

In the previous section, it was shown how the procedure selects a one-dimensional radial vector in the direction of the most uncertain boundary region. This section develops the capability to search along that vector to find the boundary point. This is shown schematically in Figure III.E-1. First, the direction is specified by a vector ρ, with the focalpoint at θ^{Des}. Then, the algorithm searches along this vector in a one-dimensional line-search until it converges to the boundary point. Thus, the search path can be represented by the equation:

$$\theta = \theta^{Des} + \delta\rho \qquad \text{Eq III.E-1}$$

where δ is a positive constant to be varied until convergence, shown by $\hat{\delta}$. $\mathbf{g}$ is negative at the design point, increasing until it reaches zero at the boundary point.

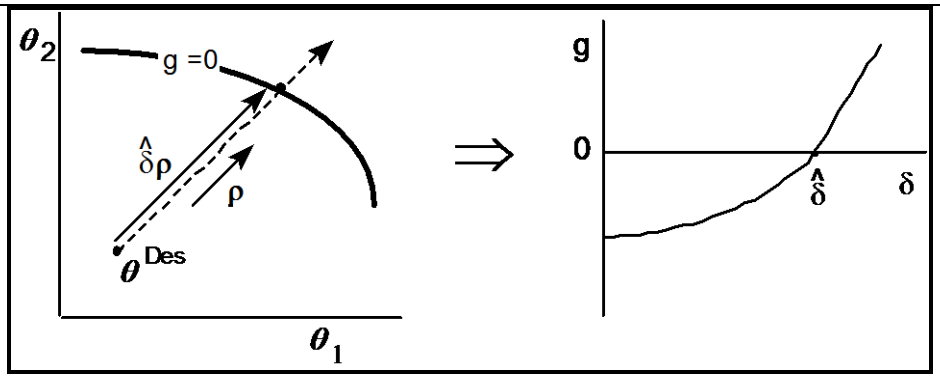


Figure III.E-1
One-Dimensional Line-Search to Find Boundary Point in Any Specified Direction

The line-search procedure has two extremes of solution methods:

Nested-Loop: Brute force approach, using nested loops to solve the process simulation model (x), optimize the control variables (z), and search for the boundary point $\hat{\theta}$.

Simultaneous : More elegant approach, which solves all three parts simultaneously.

The nested-loop algorithm requires no process simulator modification, and is simple to program. However, it is slow, requiring several converged process simulations for each

boundary point. The simultaneous algorithm requires process simulation modification, but should be faster.

The simultaneous algorithm automatically provides the gradient at each boundary point. The nested algorithm must calculate this by finite-difference approximation, which is expensive.

Another issue is convergence precision: How precisely must we calculate the location of the boundary points? Naturally, a looser convergence tolerance requires less simulation effort. The nested-loop algorithm handles this in a straightforward manner, since we can easily obtain upper and lower bounds along the radial vector. However, since the simultaneous method is converging the process model while it is searching along the radial direction, it is more difficult to bound the boundary point and quantify the precision.

Finally, the presence of control variables complicates both approaches. For the nested-loop algorithm, a third level of nested iterations is required, significantly increasing the computational effort. For the simultaneous algorithm, extra simulation effort is required. However, it is less severe.

Either algorithm is acceptable, since they both converge to the same answer. For geometric testing, the simultaneous approach was used to prove the algorithm. However, for distillation testing, the nested-loop algorithm was used.

III.E.1. Nested-loop Mathematical Algorithm

This section develops the nested-loop algorithm. It was used for distillation testing. The algorithm is presented including control variables, but these are not required for distillation.

The nested-loop algorithm separates the boundary search and the process simulation into nested loops as shown in Figure III.E.1-1. The inner loop converges the process model using the process simulator. The outer loop converges the one-dimensional line-search using a procedure such as bisection or regula-falsi. When control variables are present, a third, intermediate loop is required to optimize the control variables.

Two options for the outer loop (boundary search) of the nested-loop algorithm are:

1) Start at a relatively short distance, then step outwards until the boundary is encountered.
2) Start at some maximum distance, then step back if the boundary was encountered.

The problem with the first approach is that the boundary may not even exist in certain directions, i.e., is open-ended. Or the search may lead into regions that are extremely difficult to converge in the inner loop (process model simulation) because of unrealistic parameter values. Consequently, too much effort is spent in statistically insignificant regions.

The second approach avoids this problem. The maximum distance is chosen to be sufficiently large so the region beyond it is statistically insignificant. The line-search begins at this point. If this point is found to be outside the constraint boundary, the algorithm steps inward and converges. If this point is inside the constraint boundary (meaning the constraint boundary is still further out), the search is halted and the maximum distance point becomes the boundary point (called a *nonfailure* point since it is not actually on the boundary).

The flowchart for the outer loop is shown in Figure III.E.1-2. The actual convergence algorithm and convergence criterion are not included.

The one-dimensional line-search function may be highly nonlinear, sometimes with first-order discontinuity as the active constraint switches. A variety of convergence algorithms can be used, e.g., bisection and linear or quadratic interpolation. Bisection is the slowest method, but is very stable and hence useful for the early iteration stages. Linear interpolation by regula-falsi is more powerful, but has difficulty with highly nonlinear situations. Also, regula-falsi does not keep track of upper and lower bounds and simply interpolates from the two most recent points, which can lead to oscillations and

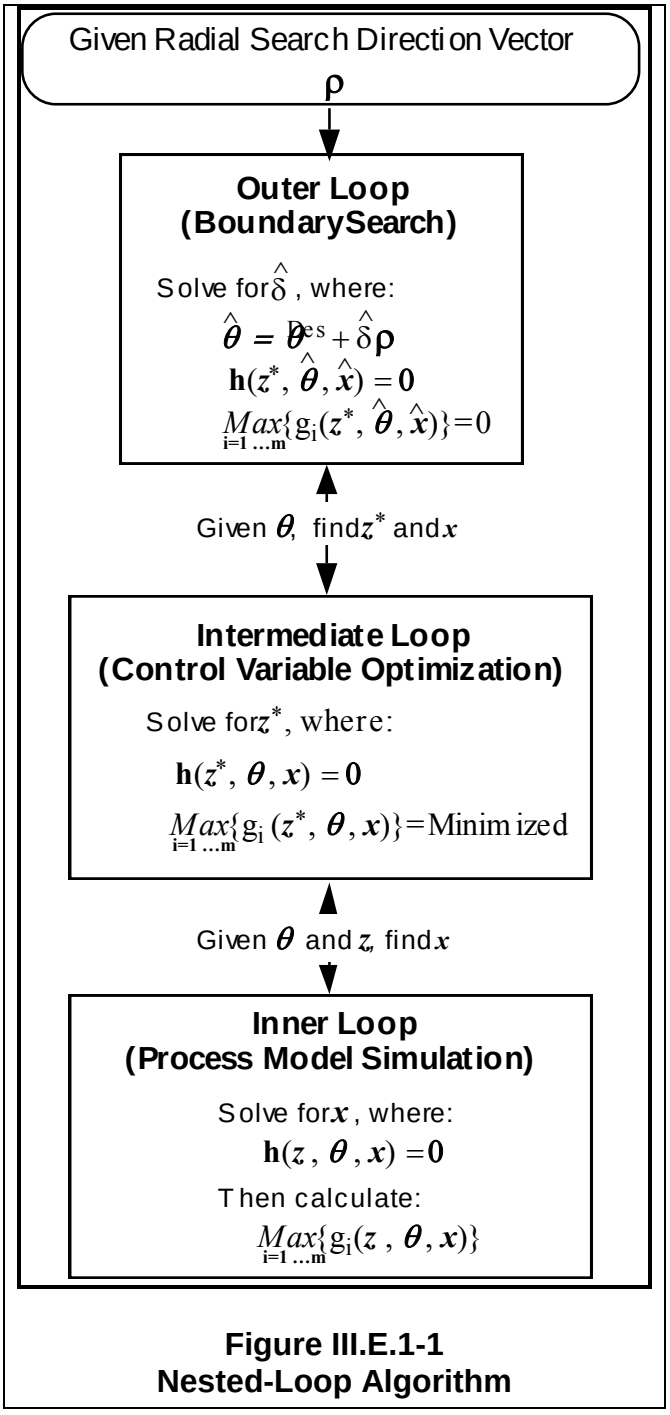


Figure III.E.1-1
Nested-Loop Algorithm

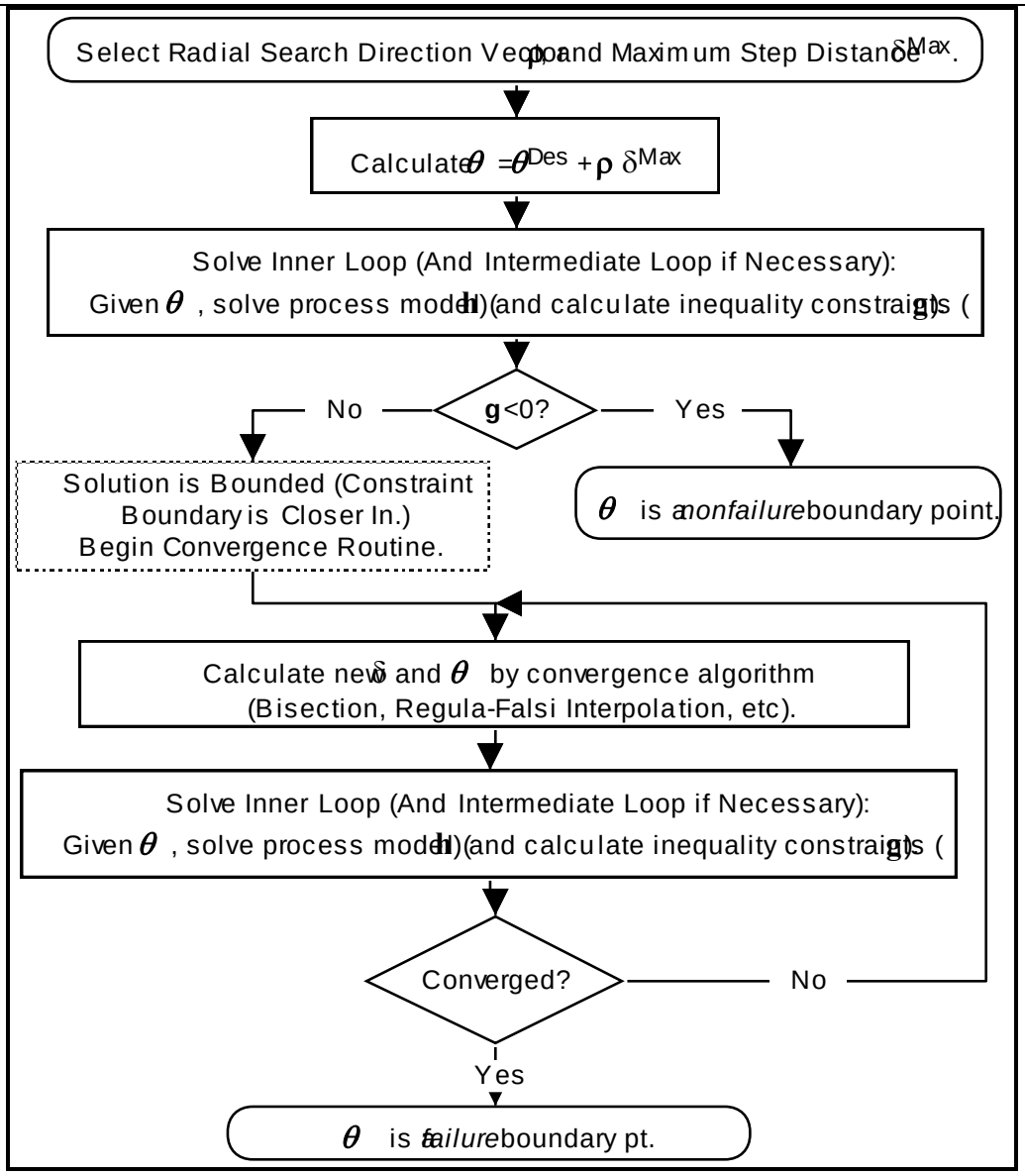


Figure III.E.1-2. Schematic of Outer-loop Boundary-Search Algorithm

nonconvergence. This can be avoided by maintaining upper and lower interpolation bounds. However, this may result in inefficient 'inching along', as shown in Figure III.E.1-3. Here, the function has a 'bathtub' shape, where there is very little change for a long distance, then a rapid increase. Linear interpolation between the upper and lower bounds causes the lower bound to increase in very small steps.

To counteract this, a simple modification was applied. Whenever the linear interpolation selects a new point close to the upper or lower bound, this is overridden and the new point is

moved towards the middle. Mathematically, the following numbers worked well: When the new point is selected within 30% of either bound, it is moved to the center by a factor of 1.5. Thus, a new point 20% from one side becomes 30% from that side, and a new point 1% from one side becomes 1.5% from that side.

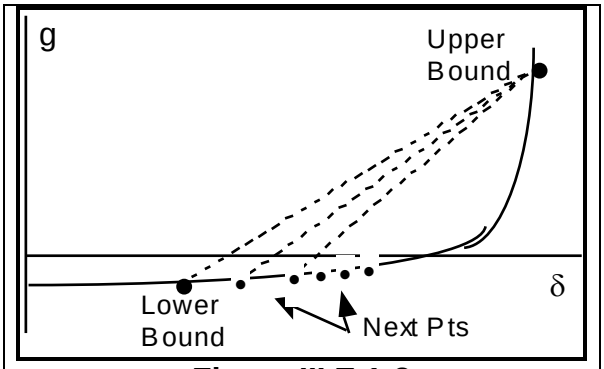


Figure III.E.1-3
Bounded Regula-Falsi Can 'Inch Along' Inefficiently

Another modification was to transform the constraint residuals (y-axis in Figure III.E.1-3) by log transformations before interpolating. This tended to linearize the curve and improve convergence.

These modifications substantially reduced the computational effort. During testing, bisection took 22 iterations to converge to the highest accuracy tolerances. The final algorithm was usually able to converge within 4-10 iterations, with 6 being standard. The first two or three points served to bisect the region, i.e., slow, but sure, reduction. Then, when the bound was sufficiently small for linearity to be accurate, a good interpolation would reduce the upper-lower bound by a factor of 100 or more.

While this algorithm worked satisfactorily, more advanced approaches may be developed to improve the computational efficiency.

III.E.1.a. Convergence Criterion

This section defines the convergence criterion for the nested-loop algorithm.

Each of the three loops of the nested-loop algorithm requires its own convergence criterion. Control variable optimization is excluded from this discussion. For the boundary search, the convergence criterion must focus on accuracy in $\hat{\delta}$. For the process model simulation, accuracy in the value of g_i is the important factor.

Two types of convergence criteria are used for the boundary search. They are:

$$\delta^{Max} - \delta^{Min} < \varepsilon_{\delta_1} \qquad \text{Eq III.E.1.a-1}$$

$$\left| \underset{i=1\ldots m}{Max}\{g_i\} \right| < \varepsilon_{\delta_2} \qquad \text{Eq III.E.1.a-2}$$

Convergence is achieved when either equation is satisfied. Eq III.E.1.a-1 indicates that the line-search has bounded the solution to a sufficiently precise value in θ. This is the most important quantity because it relates directly to the potential error in the reliability (θ-space).

Eq III.E.1.a-2 is also used to improve efficiency because the line-search often 'hits' the boundary point before the upper-lower bounds have decreased within the ε_{δ_1} criterion.

For the process model simulation inner-loop it is difficult to define convergence accuracy in g_i because the distillation process simulation model is iterative. The convergence criterion of the distillation process simulation model is presented in section IV.C. However, it is based on the iterative solution to solve for $\boldsymbol{x}$ where $\mathbf{h}(\boldsymbol{x})=\mathbf{0}$, which has no simple relation to accuracy in g_i. To avoid this problem, distillation convergence tolerances were set as tight as possible to ensure maximum accuracy in g_i. This decreases computational efficiency, however.

It is difficult to determine optimum tolerances for the convergence criteria, i.e., loosest possible values that maintain accuracy, because the nested-loops interact. If one tolerance is loose while the other is tight, performance is poor. For example, if the process simulation model convergence tolerance is loose, then the line-search interpolates with inaccurate values of g_i, which is inefficient and may cause failure. On the other hand, there is much room for optimization of the tolerances. For instance, the search may start with loose tolerances, then tighten them as the iteration proceeds. Or convergence criteria may be based on the statistical significance—looser tolerances for the less statistically significant locations.

For this research, very tight tolerances were used, because accuracy was most important. It is recommended that further studies adjust these tolerances more optimally.

III.E.2. Simultaneous Convergence Mathematical Algorithm

This section develops the simultaneous algorithm. This is more complicated than the nested-loop algorithm, but should be substantially faster. It was partially tested on geometric test problems, but was not used for distillation. Process simulator modification is required.

The simultaneous algorithm converges the process model and line-search *at the same time*. That is, it solves for $\hat{\boldsymbol{x}}$, z^* and $\hat{\boldsymbol{\theta}}$, where $\mathbf{h}(\hat{\boldsymbol{x}}, z^*, \hat{\boldsymbol{\theta}})=\mathbf{0}$, $\underset{i=1,..,m}{Max}\{g_i(\hat{\boldsymbol{x}}, z^*, \hat{\boldsymbol{\theta}})\}=0$, in one step, i.e., one converged simulation.

For distillation, the method is based on the Naphtali-Sandholm algorithm (see Section IV.A.3), which solves for $\mathbf{h}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\theta}})=\mathbf{0}$ given $\boldsymbol{\theta}$, by Newton's method, and requires finite difference evaluation of $\partial\mathbf{h}/\partial\boldsymbol{x}$. The new algorithm requires the additional calculation of the Jacobians $\partial\mathbf{h}/\partial\boldsymbol{\theta}$, $\partial\mathbf{g}/\partial\boldsymbol{x}$, $\partial\mathbf{g}/\partial\boldsymbol{\theta}$ (additional Jacobians with respect to z when present). Partial derivatives with respect to $\mathbf{g}$ should be analytically simple, but $\partial\mathbf{h}/\partial\theta$ will generally require finite differencing.

These additional Jacobians are the only major extra computational effort. Some additional matrix manipulations are required, but these are not severe.

III.E.2.a. Without Control Variables

Every point on the constraint boundary satisfies the equations:

$$\mathbf{h}(\boldsymbol{x}, \boldsymbol{\theta}) = \mathbf{0} \qquad (n) \qquad \text{Eq III.E.2.a-1}$$

$$\underset{i=1\ldots m}{Max}\left[g_i(\boldsymbol{x}, \boldsymbol{\theta}) = 0\right] \qquad (1) \qquad \text{Eq III.E.2.a-2}$$

Eq III.E.2.a-1 is the process model solution. Eq III.E.2.a-2 is the boundary condition.

There are $n+1$ equations in $n+p$ unknowns, resulting in $p-1$ degrees of freedom. Therefore, solution will ultimately result in one equation in p variables for the boundary:

$$f(\boldsymbol{\theta}) = 0 \qquad (1) \qquad \text{Eq III.E.2.a-3}$$

First, Eq III.E.2.a-1 is linearized to get:

$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{x}}\Delta\boldsymbol{x} + \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}\Delta\boldsymbol{\theta} \;=\; \Delta\mathbf{h} \qquad (n) \qquad \text{Eq III.E.2.a-4}$$

Eq III.E.2.a-2 cannot be linearized directly, because of the *Max*() condition. However, there is another way. First, we expand *all* the inequality constraints:

$$\frac{\partial \mathbf{g}}{\partial \boldsymbol{x}}\Delta\boldsymbol{x} + \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}\Delta\boldsymbol{\theta} \;=\; \Delta\mathbf{g} \qquad (m) \qquad \text{Eq III.E.2.a-5}$$

We now wish to extrapolate $\mathbf{h}$ to zero. Anticipating an iterative solution, use superscripts k and $k+1$ for the iteration number:

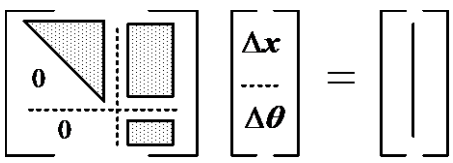
$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{x}}\left(\boldsymbol{x}^{k+1} - \boldsymbol{x}^{k}\right) + \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}\left(\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^{k}\right) \;=\; -\mathbf{h}^{k} \qquad (n) \qquad \text{Eq III.E.2.a-6}$$

$$\frac{\partial \mathbf{g}}{\partial \boldsymbol{x}}\left(\boldsymbol{x}^{k+1} - \boldsymbol{x}^{k}\right) + \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}\left(\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^{k}\right) \;=\; -\mathbf{g}^{k} \qquad (m) \qquad \text{Eq III.E.2.a-7}$$

which can be written in matrix form (dropping the superscripts for clarity):

$$\begin{bmatrix} \dfrac{\partial \mathbf{h}}{\partial \boldsymbol{x}} & \dfrac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \\ \dfrac{\partial \mathbf{g}}{\partial \boldsymbol{x}} & \dfrac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} \end{bmatrix} \begin{bmatrix} \Delta\boldsymbol{x} \\ \Delta\boldsymbol{\theta} \end{bmatrix} \;=\; \begin{bmatrix} -\mathbf{h} \\ -\mathbf{g} \end{bmatrix} \qquad \text{Eq III.E.2.a-8}$$

Solution now proceeds by the forward step of Gaussian elimination. Starting from the top, each row is normalized by its diagonal in $\partial\mathbf{h}/\partial\boldsymbol{x}$, and all columns below are subtracted out. This continues through each row in $\partial\mathbf{h}/\partial\boldsymbol{x}$ to give:



Eq III.E.2.a-9

where the lines, boxes and triangles indicate nonzero components.

The bottom m rows now provide m distinct equations:

$$\mathbf{A}_i\left(\theta^{k+1} - \theta^k\right) = \mathbf{B}_i \qquad i = 1\ldots m \qquad \text{Eq III.E.2.a-10}$$

where $\mathbf{A}$ and $\mathbf{B}$ are constants of dimension $[1 \cdot p]$ and $[1 \cdot 1]$ respectively.

Each of the m equations represents the linearized boundary of inequality constraint i, $i=1\ldots m$. This is the linearized form of Eq III.E.2.a-3. However, only one of these equations is active. This is the one that fails nearest the focalpoint (θ^{Des}), see Figure III.E.2.a-1. The other $m\text{-}1$ equations are 'farther out' or 'behind' the active constraint and can be ignored.

We wish to step to the active constraint along the specified radial direction ρ. This can be solved for by combining Eq III.E.2.a-10 with the following equation for the radial vector:

$$\theta^{k+1} - \theta^k = \mathbf{C} \cdot \rho$$

where $\mathbf{C}$ is a [1·1] constant.

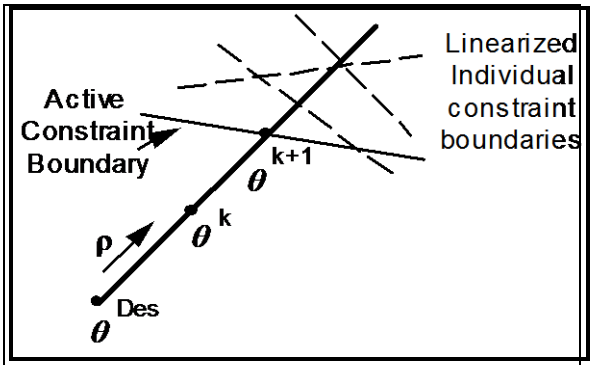


Figure III.E.2.a-1
Iteration Step in Radial Direction

We substitute Eq III.E.2.a-11 into each of the m equations Eq III.E.2.a-10 to give:

$$\mathbf{C}_i \cdot \mathbf{A}_i \cdot \rho = B_i \qquad i = 1\ldots m \qquad \text{Eq III.E.2.a-12}$$

where $\mathbf{C}_i$ denotes the $\mathbf{C}$ constant for the i'th equation. This may be written:

$$\mathbf{C}_i = \frac{B_i}{\sum_{j=1}^{p} \mathbf{A}_{i_j} \cdot \rho_j} \qquad i = 1\ldots m \qquad \text{Eq III.E.2.a-13}$$

where

$\mathbf{A}_{i_j}$ is the j'th column of vector $\mathbf{A}_i$

$\mathbf{C}_i$ represents the step distance along the radial vector from θ^k. The active constraint is the one with the smallest $\mathbf{C}_i$, or if the boundary has been passed, the most negative. That is, $\mathbf{C}^{Min} = \underset{i=1\ldots m}{Min}\left(\mathbf{C}_i\right)$. Then:

$$\theta^{k+1} = \theta^k + \mathbf{C}^{Min} \cdot \rho \qquad \text{Eq III.E.2.a-14}$$

Eq III.E.2.a-14 provides the new iteration point $\boldsymbol{\theta}^{k+1}$. However, we still need to solve for $\boldsymbol{x}^{k+1}$. This is done by completing the matrix inversion Eq III.E.2.a-9. The $\Delta\boldsymbol{\theta}$ in this matrix is $\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^{k}$, which was just calculated. Inserting this into the equation eliminates the remaining degrees of freedom and gives:

$$\begin{bmatrix} \nabla\mathbf{h}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \Delta\boldsymbol{x} \\ \hline \Delta\boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \hline \mathbf{h} \end{bmatrix} \qquad \text{Eq III.E.2.a-15}$$

$\Delta\boldsymbol{x}$ (and hence $\boldsymbol{x}^{k+1}$) can be solved for immediately by backwards Gaussian elimination, stepping back up the matrix.

This completes one iteration. We now re-linearize and repeat the iteration until convergence, where $\mathbf{h} = \mathbf{0}$ and one of the g_i equals zero, with the others negative.

III.E.2.a.i. Orthogonal Boundary Point

An orthogonal point is the point on each constraint boundary 'nearest' the focalpoint $\boldsymbol{\theta}^{Des}$, see Figure III.E.2.a.i-1. This will be denoted $\hat{\boldsymbol{\theta}}^{\boldsymbol{i}}$, where superscript ^{i} denotes the i'th inequality constraint. These points might be especially useful as 'key' points in accurately approximating the boundary.

Orthogonal points occur where the gradient is parallel to the radial direction. The nested-loop algorithm does not calculate the gradient until after convergence, so this algorithm cannot provide orthogonal points without great effort. However, the simultaneous algorithm calculates the gradient throughout the iteration, so this is not a problem.

The orthogonal violation point $\hat{\boldsymbol{\theta}}^{\boldsymbol{i}}$ (and every other point on the boundary of constraint g_i) satisfies the equations:

$$\mathbf{h}(\boldsymbol{x}, \boldsymbol{\theta}) = \mathbf{0}$$

$$g_i(\boldsymbol{x}, \boldsymbol{\theta}) = 0$$

First, Eqs III.E.2.a.i-1 and III.E.2.a.i-2 are linearized to get:

$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{x}} \Delta\boldsymbol{x} + \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \Delta\boldsymbol{\theta} = \Delta\mathbf{h}$$

$$\frac{\partial g_i}{\partial \boldsymbol{x}} \Delta\boldsymbol{x} + \frac{\partial g_i}{\partial \boldsymbol{\theta}} \Delta\boldsymbol{\theta} = \Delta g_i$$

Extrapolating to zero, and writing in iterative form:

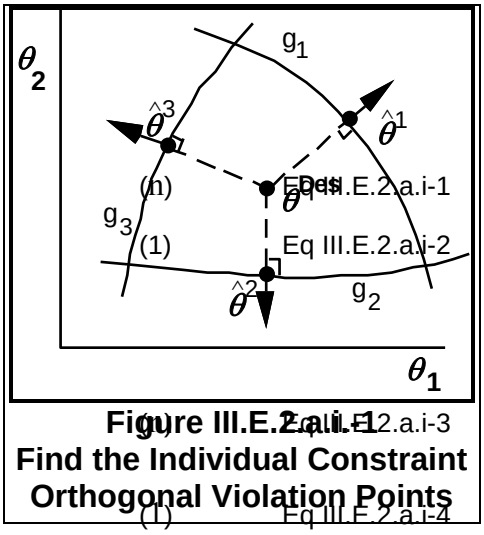


Figure III.E.2.a.i-1 Find the Individual Constraint Orthogonal Violation Points

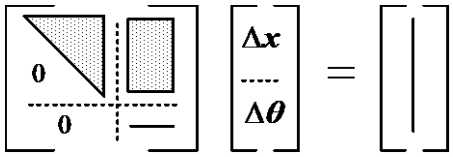
$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{x}}\left(\boldsymbol{x}^{k+1}-\boldsymbol{x}^{k}\right)+\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}\left(\boldsymbol{\theta}^{k+1}-\boldsymbol{\theta}^{k}\right) \;=\; -\mathbf{h}^{k} \qquad \text{(n)} \qquad \text{Eq III.E.2.a.i-5}$$

$$\frac{\partial \mathrm{g_i}}{\partial \boldsymbol{x}}\left(\boldsymbol{x}^{k+1}-\boldsymbol{x}^{k}\right)+\frac{\partial \mathrm{g_i}}{\partial \boldsymbol{\theta}}\left(\boldsymbol{\theta}^{k+1}-\boldsymbol{\theta}^{k}\right) \;=\; -\mathrm{g_i^{k}} \qquad \text{(1)} \qquad \text{Eq III.E.2.a.i-6}$$

which can be written in matrix form (dropping the superscripts):

$$\begin{bmatrix} \dfrac{\partial \mathbf{h}}{\partial \boldsymbol{x}} & \dfrac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \\ \dfrac{\partial \mathrm{g_i}}{\partial \boldsymbol{x}} & \dfrac{\partial \mathrm{g_i}}{\partial \boldsymbol{\theta}} \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{x} \\ \Delta \boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} -\mathbf{h} \\ -\mathrm{g_i} \end{bmatrix} \qquad \text{Eq III.E.2.a.i-7}$$

Solution now proceeds by forward Gaussian elimination. Starting from the top, each row is normalized by its diagonal in $\partial \mathbf{h}/\partial \boldsymbol{x}$, and all rows below are subtracted out. This continues through all the rows in $\partial \mathbf{h}/\partial \boldsymbol{x}$ to give:



Eq III.E.2.a.i-8

The bottom row now provides the following information:

$$\mathbf{A}\left(\boldsymbol{\theta}^{k+1}-\boldsymbol{\theta}^{k}\right)=\mathbf{B} \qquad \text{Eq III.E.2.a.i-9}$$

where $\mathbf{A}$ and $\mathbf{B}$ are constants of dimension $[1 \cdot \mathrm{p}]$ and $[1 \cdot 1]$ respectively. So we have 1 equation in p unknowns.

Eq III.E.2.a.i-9 is the linearized approximation to the constraint boundary surface. To solve the nonlinear problem, we pick one point on the linearized line, re-linearize at this point, and repeat until convergence. We choose the point on the linearized line closest to the focalpoint (θ^{Des}), which is equivalent to the condition of orthogonality, see Figure III.E.2.a.i-2.

θ^{k+1} must lie on the surface $\mathbf{A}\left(\theta^{k+1}-\theta^{k}\right)=\mathbf{B}$. Also, the line $\theta^{k+1}-\theta^{Des}$ is perpendicular to the surface, so we can write a point-slope equation for this line as $\theta^{k+1}=\theta^{Des}+\mathbf{C}\cdot\mathbf{A}^{\mathbf{T}}$, where $\mathbf{C}$ is a [1·1] constant. Thus, we have:

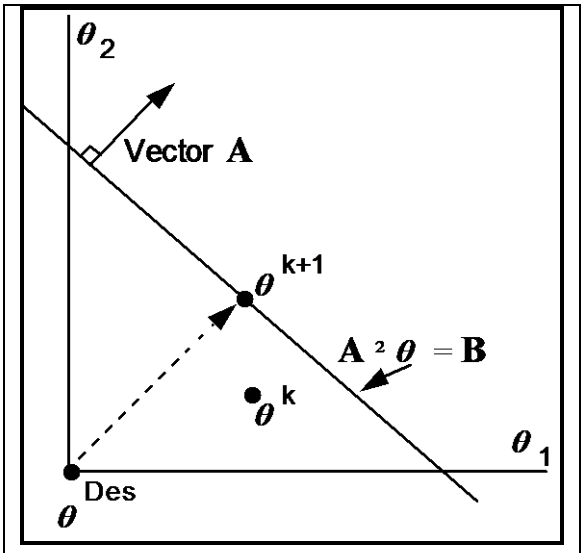


Figure III.E.2.a.i-2
Step to Linearized Boundary that Minimizes Distance from Design Point

$$\mathbf{A}\left(\theta^{k+1}-\theta^{k}\right)=\mathbf{B} \qquad \text{Eq III.E.2.a.i-10}$$

$$\theta^{k+1}=\theta^{Des}+\mathbf{C}\cdot\mathbf{A}^{\mathbf{T}} \qquad \text{Eq III.E.2.a.i-11}$$

$$\theta^{k+1}-\theta^{Des} = \left(\theta^{k+1}-\theta^{k}\right)+\left(\theta^{k}-\theta^{Des}\right) \qquad \text{Eq III.E.2.a.i-12}$$

which can be combined to solve for θ^{k+1}:

$$\theta^{k+1}-\theta^{Des} = \frac{\mathbf{A}^{\mathbf{T}}\left\{\mathbf{B}+\mathbf{A}\left(\theta^{k}-\theta^{Des}\right)\right\}}{\mathbf{A}\cdot\mathbf{A}^{\mathbf{T}}} \qquad \text{Eq III.E.2.a.i-13}$$

Eq III.E.2.a.i-13 provides the new iteration point θ^{k+1}. Solving for x^{k+1} is done by completing the matrix inversion as before, using backwards Gaussian elimination.

This completes the linearized iteration. Now the nonlinear model is re-evaluated at x^{k+1}, θ^{k+1}, the process is re-linearized, and the iteration is repeated until convergence which occurs when $\mathbf{h}$ and g_i are zero, and x and θ are mapped onto themselves. This orthogonal point is denoted $\hat{x}^{i}, \hat{\theta}^{i}$.

Finally, the other inequality constraints are checked to make sure they are inactive (i.e., determine if $g_j\left(\hat{x}^{i}, \hat{\theta}^{i}\right)\le 0 \quad \text{for all } j=1\ldots m, j\neq i$). If any other constraint is active, the orthogonal point is not useful.

Confirmation of orthogonality may be done as follows: The vector $(\hat{\boldsymbol{\theta}}^{\boldsymbol{i}} - \boldsymbol{\theta}^{\boldsymbol{Des}})$ is parallel to the gradient of g_i in the "constrained space" where $\mathbf{h}=\mathbf{0}$. So to the first-order approximation:

$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{x}} \Delta \hat{\boldsymbol{x}}^{\boldsymbol{i}} + \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \Delta \hat{\boldsymbol{\theta}}^{\boldsymbol{i}} = \mathbf{0} \qquad \text{(n)} \qquad \text{Eq III.E.2.a.i-14}$$

$$\frac{\partial g_i}{\partial \boldsymbol{x}} \Delta \hat{\boldsymbol{x}}^{\boldsymbol{i}} + \frac{\partial g_i}{\partial \boldsymbol{\theta}} \Delta \hat{\boldsymbol{\theta}}^{\boldsymbol{i}} = 0 \qquad \text{(1)} \qquad \text{Eq III.E.2.a.i-15}$$

From III.E.2.a.i-14:

$$\Delta \hat{\boldsymbol{x}}^{\boldsymbol{i}} = -\left(\frac{\partial \mathbf{h}}{\partial \boldsymbol{x}}\right)^{-1} \left[\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \Delta \hat{\boldsymbol{\theta}}^{\boldsymbol{i}}\right] \qquad \text{Eq III.E.2.a.i-16}$$

and inserting this into Eq III.E.2.a.i-15 gives:

$$\left[\frac{\partial g_i}{\partial \boldsymbol{\theta}} - \frac{\partial g_i}{\partial \boldsymbol{x}} \left(\frac{\partial \mathbf{h}}{\partial \boldsymbol{x}}\right)^{-1} \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}\right] \Delta \hat{\boldsymbol{\theta}}^{\boldsymbol{i}} = 0 \qquad \text{Eq III.E.2.a.i-17}$$

Now the vector inside the square brackets of Eq III.E.2.a.i-17 is the gradient vector of g_i with respect to $\boldsymbol{\theta}$, such that for whatever variation in $\hat{\boldsymbol{\theta}}^{\boldsymbol{i}}$, $\hat{\boldsymbol{x}}^{\boldsymbol{i}}$ is varied to maintain $\mathbf{h}=\mathbf{0}$ (at least to the linear approximation). Thus, the angle between this vector and $(\hat{\boldsymbol{\theta}}^{\boldsymbol{i}} - \boldsymbol{\theta}^{\boldsymbol{Des}})$ is zero at convergence. Note that this angle makes a useful iterative observation variable to track convergence.

The vector inside the square brackets of Eq III.E.2.a.i-17 is also the gradient vector that the boundary-approximation procedure requires. This is true for radial points as well as the orthogonal points.

III.E.2.b. With Control Variables

This is a more difficult problem, because the control variables $\boldsymbol{z}$ must be optimized simultaneously. This has the effect of 'pushing back' the constraint boundary. The optimum $\boldsymbol{z}$ occurs when either $\partial g_i / \partial \boldsymbol{z} = 0$, or $\boldsymbol{z}$ is at its maximum or minimum settings. As Kubic and Stein (1988) have noted, the latter situation occurs for many practical situations. However, for the general situation, the optimum $\boldsymbol{z}$, denoted $\boldsymbol{z}^*$, is unknown. Then a search is required.

The equations that all boundary points satisfy are:

$$\mathbf{h}\left(\boldsymbol{z}^*, \boldsymbol{x}, \boldsymbol{\theta}\right) = \mathbf{0} \qquad \text{(n)} \qquad \text{Eq III.E.2.b-1}$$

$$\underset{\boldsymbol{z}^* \in \mathbf{T}}{Min} \left\{ \underset{i=1,\ldots m}{Max} \left[g_i\left(\boldsymbol{z}^*, \boldsymbol{x}, \boldsymbol{\theta}\right) \right] = 0 \right\} \qquad \text{(1)} \qquad \text{Eq III.E.2.b-2}$$

where $\mathbf{T}$ is the range of all possible control variable settings, e.g.:

$$\left(\mathbf{T} \mid z_j^{Min} < z_j^* < z_j^{Max}\right) \quad j = 1 \ldots q \qquad \text{Eq III.E.2.b-3}$$

Eq III.E.2.b-1 is the process simulation model (equality constraints), with control variables included. The $Max\{\}$ of Eq III.E.2.b-2 is the boundary condition (the *active* inequality constraint is zero). The $Min\{\}$ of Eq III.E.2.b-2 is the control variable optimization.

Thus we now have $n+1$ equations in $n+p+q$ unknowns. But, there are still only $p-1$ degrees of freedom, because z^* is fixed for every θ.

An alternative formulation is to write first order derivatives in z, as follows. First, for Nomenclature and equation simplicity, use g' to denote the active constraint $\underset{i=1\ldots m}{Max}\left[g_i\left(z^*, x, \theta\right)\right]$. And temporarily ignore the difficulty of determining *which* constraint is active. Eq III.E.2.b-2 is equivalent to:

$$g'\left(z^*, x, \theta\right) = 0 \qquad \text{Eq III.E.2.b-4}$$

and

$$\frac{\partial g'}{\partial z_j^*} = 0 \qquad \left(z_j^* \mid z_j^{Min} < z_j^* < z_j^{Max}\right) \qquad \text{Eq III.E.2.b-5a}$$

or

$$\frac{\partial g'}{\partial z_j^*} > 0 \qquad \left(z_j^* \mid z_j^* = z_j^{Min}\right) \qquad \text{Eq III.E.2.b-5b}$$

or

$$\frac{\partial g'}{\partial z_j^*} < 0 \qquad \left(z_j^* \mid z_j^* = z_j^{Max}\right) \qquad \text{Eq III.E.2.b-5c}$$

This converts the problem into $n+q+1$ equations. We can linearize Eqs III.E.2.b-1, 4 and 5a to get:

$$\frac{\partial \mathbf{h}}{\partial z}\Delta z + \frac{\partial \mathbf{h}}{\partial x}\Delta x + \frac{\partial \mathbf{h}}{\partial \theta}\Delta\theta = \Delta \mathbf{h} \qquad \text{Eq III.E.2.b-6}$$

$$\frac{\partial g'}{\partial z}\Delta z + \frac{\partial g'}{\partial x}\Delta x + \frac{\partial g'}{\partial \theta}\Delta\theta = \Delta g' \qquad \text{Eq III.E.2.b-7}$$

$$\frac{\partial^2 g'}{\partial z^2}\Delta z = -\frac{\partial g'}{\partial z} \qquad \text{Eq III.E.2.b-8}$$

Eq III.E.2.b-8 appears to offer a simple approach to obtain Δz, by the second-order equivalent of Newton's method: Calculate $\partial g'/\partial z$ and $\partial^2 g'/\partial z^2$, then solve Eq III.E.2.b-8 for Δz. Then, holding the new z constant, solve for $\Delta\theta$, Δx as before. Unfortunately, this will not work because g' is likely to be an implicit function of z connected only through $\mathbf{h}=\mathbf{0}$.

That is, z might not appear in g', so all direct calculations of the partial derivatives will be zero. Therefore, a different approach is needed.

There are two iterative options: (1) One is a simultaneous approach, in which both θ and z (and $\boldsymbol{x}$ of course) are varied during each iteration. (2) The other is a two-step alternating approach: A) Hold θ constant and step in z. B) Then hold z constant and step in θ. Repeat A and B again until convergence.

The two-step alternating approach has the advantage that the latter step has already been developed in the previous section. The z iteration step is now developed:

Holding θ constant, a first-order expansion of $\mathbf{h}$ and g' gives:

$$\frac{\partial \mathbf{h}}{\partial z}\Delta z + \frac{\partial \mathbf{h}}{\partial \boldsymbol{x}}\Delta \boldsymbol{x} = \Delta \mathbf{h} \qquad \text{Eq III.E.2.b-9}$$

$$\frac{\partial g'}{\partial z}\Delta z + \frac{\partial g'}{\partial \boldsymbol{x}}\Delta \boldsymbol{x} = \Delta g' \qquad \text{Eq III.E.2.b-10}$$

Taking the current point as iteration k, and extrapolating Eq III.E.2.b-9 to zero, we can eliminate $\Delta \boldsymbol{x}$:

$$\Delta \boldsymbol{x} = -\left(\frac{\partial \mathbf{h}}{\partial \boldsymbol{x}}\right)^{-1}\left(\mathbf{h}^k + \frac{\partial \mathbf{h}}{\partial z}\Delta z\right) \qquad \text{Eq III.E.2.b-11}$$

and substitute this into Eq III.E.2.b-10 to give (with some rearrangement):

$$\left[\frac{\partial g'}{\partial z} - \frac{\partial g'}{\partial \boldsymbol{x}}\left(\frac{\partial \mathbf{h}}{\partial \boldsymbol{x}}\right)^{-1}\frac{\partial \mathbf{h}}{\partial z}\right]\Delta z = \Delta g' + \frac{\partial g'}{\partial \boldsymbol{x}}\left(\frac{\partial \mathbf{h}}{\partial \boldsymbol{x}}\right)^{-1}\mathbf{h}^k \qquad \text{Eq III.E.2.b-12}$$

The matrix inside the square parentheses in Eq III.E.2.b-12 is the gradient in z, which gives the direction of steepest descent in g' as a function of z, subject to satisfying the linearized process model $\mathbf{h}=\mathbf{0}$. Thus, z should be varied in this direction.

There is no prior knowledge of the optimum step distance for z, so a simple line-search must suffice to minimize g'. However, this step is only as accurate as the linear approximation in $\mathbf{h}$. Therefore, during the first few iterations when the boundary point solution is still far away, the minimum point in the line-search is not essential. As convergence approaches and the linearization becomes more accurate, the line-search optimization may be tightened.

The complete two-step algorithm is shown in Table III.E.2.b-1.

Table III.E.2.b-1
Two-Step Algorithm to Find the Constraint Boundary and Optimize the Control Variables

1. Hold z constant. Perform iterative step in θ (section III.E.2.a.).

   Process Model Calculations: $\mathbf{h}$, $\mathbf{g}$, $\frac{\partial \mathbf{h}}{\partial x}$, $\frac{\partial \mathbf{g}}{\partial x}$, $\frac{\partial \mathbf{h}}{\partial \theta}$, $\frac{\partial \mathbf{g}}{\partial \theta}$

   Calculate new θ, x

   Identify active constraint g' (at least for this iteration)

2. Hold θ constant. Perform iterative step in z (section III.E.2.b.).

   2a. Calculate $\mathbf{h}$, g', $\frac{\partial \mathbf{h}}{\partial x}$, $\frac{\partial g'}{\partial x}$, $\frac{\partial \mathbf{h}}{\partial z}$, $\frac{\partial g'}{\partial z}$

   2b. Calculate gradient: $$\nabla z^* = \left[\frac{\partial g'}{\partial z} - \frac{\partial g'}{\partial x}\left(\frac{\partial \mathbf{h}}{\partial x}\right)^{-1}\frac{\partial \mathbf{h}}{\partial z}\right]$$

   2c. Perform a line-search to minimize g' as a function of z:

   Find δ such that $g'(z,x)$ is minimized, where:

$$z = z^k - \delta \nabla z^*$$

$$x = x^k - \left(\frac{\partial \mathbf{h}}{\partial x}\right)^{-1}\left[\mathbf{h}^k + \frac{\partial \mathbf{h}}{\partial z}\left(z - z^k\right)\right]$$

   Constrain z to remain within $\mathbf{T}$. Remove control variables that are optimum at their extreme values.

3. Repeat 1 and 2 until convergence, which occurs when $\mathbf{h}=\mathbf{0}$, $g'=0$, and z is at its maximum or minimum setting or its gradient (Eq III.E.2.b-12) is zero. Finally, recheck all control variables that were optimum at their extremes during the iteration. If they are not still optimum, place them back on the active list and continue iteration.

z must be constrained to remain within its bounds $\mathbf{T}$. Any individual control variable z_i extrapolated outside this range would be set to its boundary value. At the next iteration, if its partial derivative is still nonzero and of the correct sign (Eq III.E.2.b-5), this variable is optimum at its boundary value and should be ignored for the remainder. This would be confirmed at convergence.

Convergence stability can be controlled by altering the iteration sequence in several ways: (1) Dampen the iterative steps in either θ or z. (2) Perform more than one iteration in θ while holding z constant; (3) Delay the optimization of z until the θ iteration has almost converged; (4) Or even calculate x for a few iterations while holding both θ and z constant.

Note that for each overall iteration (step 1 and 2), $\partial \mathbf{h} / \partial x$ and $\partial \mathbf{g} / \partial x$ must be calculated twice. This is the computational price paid for the two-step approach. In the early stages of the iteration, when θ and x are changing significantly, this double calculation will help stability. However, during the later iteration stages, θ and z will not be changing much, so perhaps a single evaluation of $\partial \mathbf{h} / \partial x$ and $\partial \mathbf{g} / \partial x$ per overall iteration will suffice. This is equivalent to a simultaneous approach.

Also, note the extra computational effort required to handle the control variables—first order derivatives of **h** and g' with respect to z. As already noted, g' might not be a direct function of z, so many of these derivatives will be zero. Thus, $\partial \mathbf{h} / \partial z$ is the main concern.

Testing with geometric problems has confirmed the radial and orthogonal algorithms without control variables. Some preliminary testing was done with control variables, but this was not carried sufficiently far and more work is required.

IV. Distillation Process Operation

This section introduces the distillation unit operation: First, the conventional process model is discussed, with a description of the process and a review of the simulation model and solution methods. Next, this is extended to design reliability. The inequality constraints and process uncertainties are discussed. Then, it is shown how the problem of control variables can be eliminated by substitution.

Distillation is a significant test operation because of its major design uncertainties and industrial importance—3% of the total U.S. energy consumption (Fitzmorris and Mah, 1980), and a third or more of the total energy used in a chemical plant (Linnhoff et al., 1983). However, there are many remaining uncertainties. As stated by Fair (1977):

> For certain areas, we have about reached the point of diminishing returns. "But for all areas of distillation—certainly not! A processing method that requires so much capital investment, which has no competition for making separations, and which in some aspects is supported by very weak technical foundations…"

Distillation is also computationally expensive, so it provides a realistic demonstration of the practical capabilities of the mathematical procedures.

IV.A. Conventional Process Model Development

This section describes the distillation unit operation and develops the solution (equality) constraint equations. Nonequilibrium models and control variables are discussed. Finally, solution methods are reviewed.

IV.A.1. Process Description

This section presents the distillation unit operation.

Distillation is the most important and commonly used separation method. A typical distillation column is shown in Figure IV.A.1-1. A feed stream containing a mixture of chemicals is fed into the column, and two purified product streams are removed, one from each end.

Heat is used to drive the separation, with a reboiler at the bottom and a condenser at the top. A vapor stream passes up the column, while a liquid stream falls down, the two phases coming into vigorous contact either discretely on staged trays, or continuously via column packing. Since the light components predominate in the vapor phase, while the heavy components stay in the liquid phase, a small separation occurs at each stage. The more stages used, the greater the separation.

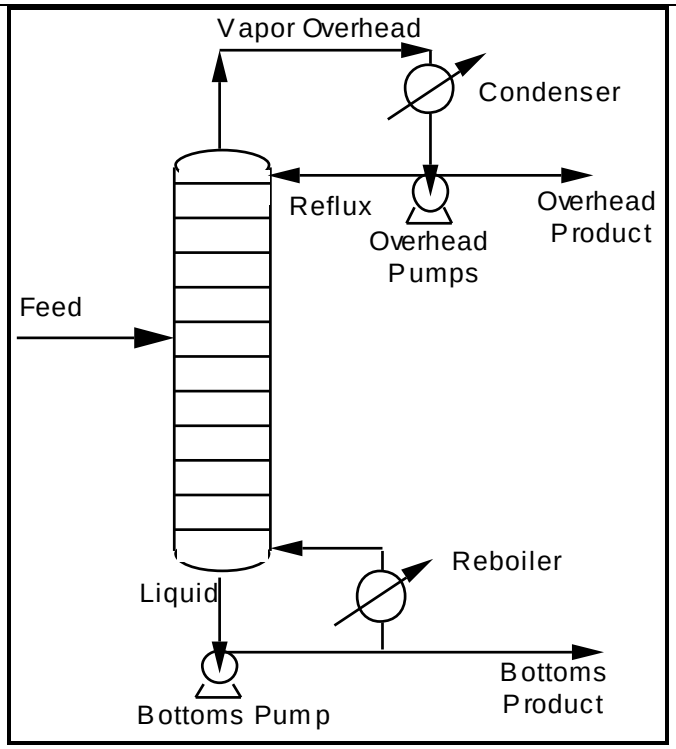


Figure IV.A.1-1
Simple Distillation Process Schematic

IV.A.2. Process Simulation Model (Equality Constraints)

This section develops the solution model (equality) equations, i.e., **h**(**x**)=0. First, a tray schematic, the mass and energy balance, and thermodynamic equilibrium equations are shown. Then, the nonequilibrium situation is discussed. Finally, control variables are introduced.

IV.A.2.a. Equilibrium Model

A schematic for a discrete stage in shown in Figure IV.A.2.a-1. Conventionally, stages are numbered from the condenser (stage 1) to the reboiler (stage N). The equations can be written in terms of total flowrates and mole fractions or component flowrates. The latter is used for this work.

The Nomenclature uses the first subscript i for the component number ($1\ldots C$), and the second subscript j for stage number ($1\ldots N$). Any stage may have a feed stream $f_{i,j}$ and/or heat addition Q_j and sidedraws U_j, W_j. All external streams (f,U,W,Q) must be specified for solution. Pressure is fixed for each stage. Typically, a designer will set the pressure at the top and assume a nominal pressure drop across each stage.

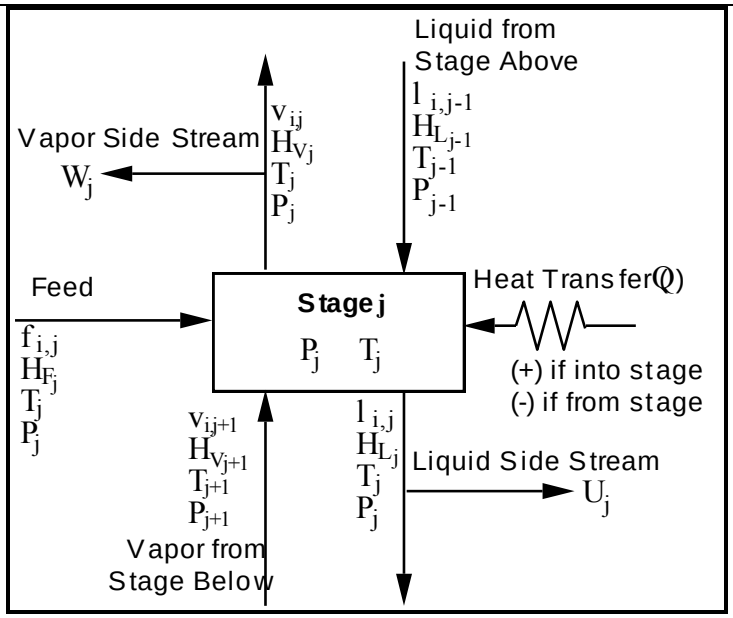


Figure IV.A.2.a-1
Distillation Discrete Stage Schematic

Table IV.A.2.a-1 lists the 2C+1 unknowns per stage, which are termed the 'state' variables. (Technically, pressure is also a state variable, however it is not being solved for.) Therefore, we need 2C+1 equations per stage to complete the problem. Component material balances give C equations, and the energy balance gives one equation. Finally, if thermodynamic equilibrium is assumed, this provides the remaining C equations. These equations are listed in Table IV.A.2.a-2.

Table IV.A.2.a-1
Discrete Stage State Variables

Vapor Component Flowrates	v_{ij}	i=1…C
Liquid Component Flowrates	l_{ij}	i=1…C
Temperature	*T*	1
Total per Stage		2C+1

Table IV.A.2.a-2
Discrete Stage Equilibrium Equations

Component Material Balances: (C)

$$l_{i,j-1} + v_{i,j+1} + f_{i,j} - \left(1 + \frac{U_j}{L_j}\right) l_{i,j} - \left(1 + \frac{W_j}{V_j}\right) v_{i,j} = 0 \qquad \text{Eq IV.A.2.a-1}$$

Energy Balance: (1)

$$L_{j-1}H_{L_{j-1}} + V_{j+1}H_{V_{j+1}} + F_jH_{F_j} - (L_j + U_j)H_{L_j} - (V_j + W_j)H_{V_j} + Q_j = 0 \qquad \text{Eq IV.A.2.a-2}$$

Thermodynamic Equilibrium: (C)

$$K_{ij}\frac{l_{ij}}{L_j} - \frac{v_{ij}}{V_j} = 0 \qquad \text{Eq IV.A.2.a-3}$$

Total Equations Per Stage: (2C+1)

with the following Nomenclature substitutions:

$$L_j = \sum_i^C l_{ij} \qquad V_j = \sum_i^C v_{ij} \qquad F_j = \sum_i^C f_{ij}$$

The energy balance and phase equilibrium equations are inexact, due to uncertainty in the enthalpy and K-values, which must be calculated by an equation of state model. These models are empirical equations fitted to limited experimental data, so the uncertainties and errors can be large.

Distillation is perhaps the unit operation most sensitive to the thermodynamic models (Hernandez et al., 1984), especially the close-boiling multi-stage separations. Small errors in the thermodynamics can lead to large errors in the design. Lashmet and Szczepanski (1974) found that 7-28% overdesign of the number of stages was necessary for a 90% confidence. Streich and Kistenmacher (1979) found that a 0.3% error in the database resulted in a 5% error in the number of trays for a C_3 splitter. And high recycle rates in the flowsheet greatly increases the sensitivity to modeling errors (Rinard, 1987).

IV.A.2.b. Nonequilibrium Model

A major uncertainty is that while the thermodynamic equations assume equilibrium operation, real equipment operates at less than thermodynamic efficiency. One option is to define tray efficiencies and replace the thermodynamic equilibrium equations with nonequilibrium tray efficiency equations. This presents the problem that tray efficiencies must be known a priori. However, these are not well known. Either they must be estimated from

previous experience with similar operations (McDaniel et al., 1970; Biddulph and Ashton, 1977; Madron, 1987; Kuncíˇr and Madron, 1990), pilot plant studies (Kastánek and Standart, 1967; Sealey, 1970; Fair et al., 1982; Hufton et al., 1988) or predicted by difficult and poorly developed empirical models (Zuiderwieg, 1982; Chan and Fair, 1984ab; Vital et al., 1984abc; Krishna et al., 1989; Prado and Fair, 1990; Rocha et al., 1991).

One unfortunate mathematical consequence of the tray efficiency models is that only C-1 component efficiencies are independent. And, for lack of better information, it is usually assumed that all components have equal tray efficiencies. However, there is considerable experimental evidence that this is not the case in general (Toor and Burchard, 1960; Biddulph et al., 1988, Medina et al.,1978, 1979ab). Efficiencies may actually vary beyond the range 0-100% (Medina et al., 1979a). In this research, equal component tray efficiencies are assumed.

The Murphree tray efficiency (Henley and Seader, 1981) is used as the nonequilibrium model in this research. It is simple to program, and does not increase computational effort. The equation is given by Eq IV.A.2.b-1, and replaces the equilibrium equation Eq IV.a.2.a-3.

$$\eta K_{ij} \frac{l_{ij}}{L_j} - \frac{v_{ij}}{V_j} + (\eta - 1)\frac{v_{ij+1}}{V_{j+1}} = 0 \qquad \text{Eq IV.A.2.b-1}$$

The Murphree tray efficiency has some theoretical limitations (Lockett, 1986). While the Hausen and Standart tray efficiencies (Medina et al., 1979b; Lockett, 1986) remove some of these limitations, they require much greater computational effort and are rarely used.

An alternative is to define mass transfer models, since the separation inefficiency is a result of mass transfer limitations anyway (Krishna and Standart, 1979; Krishnamurthy and Taylor, 1985abc; Sivasubramanian et al., 1987; Sivasubramanian and Boston, 1988; Grottoli et al., 1991). Material and energy balances are written separately for each phase, and mass transfer rate equations are written across the interface. However accurate mass transfer and diffusion coefficient models for multicomponent systems are on even shakier ground than the empirical tray efficiency equations. These models will not be considered.

IV.A.2.c. Control Variables and Degrees of Freedom

As presented, the process model requires all external streams to be specified. However, this requires specifying both Q_1 and Q_n (condenser and reboiler heat). In practice, these are bad choices because they are strongly related and we usually need to solve the problem for some other situation. So it is better to substitute alternative specifications in place of the energy balance equations on the condenser and reboiler stages.

Any choices are possible, so long as the two specifications are independent of each other and the other model equations. Product stream flowrates, reflux ratios and vapor boilup

are generally easier to converge; component compositions are harder. While it is simpler to use specifications based on the top and bottom stages, Hirose et al. (1986) and Jelínek (1988) show how to handle specifications from other stages.

Once the column is in operation, these two alternative specifications, plus column pressure, become control degrees of freedom and are varied to optimize operation. Because of a wide range of dynamics, different columns require different control methods (Brignole et al., 1985; Moore, 1985; Levien and Morari, 1987; McDonald and Palazoglu, 1987; Skogestad and Morari, 1987ab; Skogestad et al., 1990; Yang et al., 1990). However, for our purposes, it is immaterial which control variables are chosen since dynamics are not being considered. *Any* two independent specifications (in addition to column pressure) completely specify the problem—a unique setting for one set can be determined from the setting of another set.

IV.A.3. Solution Methods

This section reviews the available methods for solving the distillation simulation model. It is restricted to the conventional model, i.e., only the equality constraints: **h**(**x**)=**0**. A Naphtali-Sandholm (1971) computer algorithm is available and is used in this research.

The conventional distillation simulation model can be written as a set of equality constraints, or equations ($\mathbf{h}$). That is, given the design values ($\mathbf{d}$), parameter values (θ), column pressure and external feed streams, and two extra specifications for the control variables (z), solve for the state variables ($\boldsymbol{x}$). This is the problem:

$$\mathbf{h}(\boldsymbol{x})=\mathbf{0} \qquad \text{Eq IV.A.3-1}$$

Eq IV.A.3-1 has dimension $\mathrm{N}(2\mathrm{C}+1)$. This is a large set of highly nonlinear equations. There are basically two approaches to solving these equations (Wait and Landauro, 1988): 1) Tearing methods, where subsets of the complete set of equations are solved in sequence; and 2) Simultaneous methods, where all equations are solved simultaneously. Of the former approach, two common historical algorithms are: 1) the Bubblepoint method for wide-boiling systems, and 2) the Sum-Rates method for narrow-boiling systems. Of the simultaneous methods, Naphtali and Sandholm (1971) were the first to develop the method, based on the Newton-Raphson approach.

Tearing methods such as the bubblepoint and sum-rates methods work well for fairly ideal situations. However, they are unstable for nonideal systems. Simultaneous methods are stable for a wider range of problems and allow more flexible design specifications, but they have large memory requirements, and can be very time-consuming.

The Naphtali and Sandholm (1971) simultaneous method was used for the research. One advantage for the design reliability problem is that it automatically provides partial derivatives of the process model equations with respect to the variables. These values are required for the simultaneous line-search algorithm in the boundary-approximation procedure.

Calculating these partial derivatives is the chief computational bottleneck for the simultaneous methods. The enthalpies ($\mathbf{H}$) and K-values ($\mathbf{K}$) are calculated from complex equations of state. The partial derivatives are too complicated to be calculated analytically and must be estimated by finite difference simulation. Lucia and Westman (1984) estimate that 75-90% of the calculation effort is spent estimating physical properties and their derivatives.

One way to reduce this computational effort by avoiding finite difference is the hybrid Newton approach developed by Lucia and co-workers (Lucia and Macchietto, 1983; Lucia and Westman, 1984; Westman et al., 1984; Lucia et al., 1985; Miller and Lucia, 1985; Lucia, 1985; Venkataraman and Lucia, 1986, 1987, 1988; Kumar and Lucia, 1987; Lucia and Xu, 1988). This uses a modification of the Broyden (1965) update to estimate the partial derivatives from the iteration information. Most recently, Sridhar and Lucia (1990) have reverted to a sum-rates procedure modified by Newton's method acceleration. It is not clear if this constitutes a repudiation of the hybrid Newton methods.

Michelsen and Mollerup (1986) and Macchietto et al. (1988) recommend that all databases be modified to calculate analytically-exact partial derivatives as well as the physical property values. This can be much cheaper than finite difference methods.

Good initial estimates of the state variables ($\boldsymbol{x}$) are required for all these methods. One can use shortcut techniques, such as the classical Underwood-Fenske-Gilliland (UFG) method (see discussion in Henley and Seader, 1981, Ch12). Procedures for more complicated columns (e.g., sidestreams, azeotropic distillation) are also available (Chou and Yaws, 1988; Julka and Doherty, 1988; Swaney, 1988; Venkataraman and Lucia, 1988; Wachter and Ko, 1988; Barbosa and Doherty, 1988ab; Pham et al., 1989).

Sometimes the normal methods cannot converge for highly nonideal chemical systems. Homotopy or continuation techniques have been used (Wayburn and Seader, 1984; Hlavácek and van Rompay, 1985; Ellis et al., 1986; Vickery and Taylor, 1986; Schwartz and Stewart, 1987). These methods are guaranteed to converge, however they are very slow and require separate programs. A better approach is that of Vickery et al. (1988), who realized that nonlinearity can be controlled by the value of the tray efficiency, since the phase equilibrium equations are the primary source of nonlinearity. Thus the model can be easily converged at low tray efficiencies, then tray efficiency can be incrementally increased, solving at each step.

IV.A.4. Distillation Convergence Criterion

My distillation simulator calculates residual errors separately for the: (1) material balance, (2) energy balance, (3) phase equilibrium balance, (4) top and bottom separation specifications. This separation is useful for convergence tracking and diagnosis. Residual

values are displayed and converged on an: (a) absolute, or (b) normalized/percentage basis. The former is easier to converge, but the latter is better when trace components are present and their compositions are critical.

The user must specify the convergence criterion ε_{Dist}. The routine converges when:

$$Max\left\{\varepsilon_{MB}, \varepsilon_{EB}, \varepsilon_{PE}, \varepsilon_{Top}, \varepsilon_{Btm}\right\} < \varepsilon_{Dist} \qquad \text{Eq IV.A.4-1}$$

Material Balance:

$$l_{i,j-1} + v_{i,j+1} + f_{i,j} - \left(1 + \frac{U_j}{L_j}\right) l_{i,j} - \left(1 + \frac{W_j}{V_j}\right) v_{i,j} = \varepsilon_{MB_{ij}} \qquad \text{Eq IV.A.4-2}$$

$$\varepsilon_{MB}^{abs} = \frac{1}{C}\frac{1}{N}\sum_{i}^{C}\sum_{j}^{N}\left|\varepsilon_{MB_{ij}}\right| \qquad \text{Eq IV.A.4-3a}$$

$$\varepsilon_{MB}^{pct} = \frac{1}{C}\frac{1}{N}\sum_{i}^{C}\sum_{j}^{N}\left|\frac{\varepsilon_{MB_{ij}}}{l_{ij} + v_{ij}}\right| \qquad \text{Eq IV.A.4-3b}$$

Energy Balance:

$$L_{j-1}H_{L_{j-1}} + V_{j+1}H_{V_{j+1}} + F_jH_{F_j} - \left(L_j + U_j\right)H_{L_j} - \left(V_j + W_j\right)H_{V_j} + Q_j = \varepsilon_{EB_j} \qquad \text{Eq IV.A.4-4}$$

$$\varepsilon_{EB} = \frac{10^{-4}}{N}\sum_{j}^{N}\left|\frac{\varepsilon_{EB_j}}{L_j + V_j}\right| \qquad \text{Eq IV.A.4-5}$$

(10^{-4} compensates for H units in Btu/lbm)

Phase Equilibria (Murphree Vapor Efficiency):

$$\eta K_{ij}\frac{l_{ij}}{L_j} - \frac{v_{ij}}{V_j} + (\eta - 1)\frac{v_{ij+1}}{V_{j+1}} = \varepsilon_{PE_{ij}} \qquad \text{Eq IV.A.4-6}$$

$$\varepsilon_{PE}^{abs} = \frac{1}{C}\frac{1}{N}\sum_{i}^{C}\sum_{j}^{N}\left|\varepsilon_{PE_{ij}}\right| \qquad \text{Eq IV.A.4-7a}$$

$$\varepsilon_{PE}^{pct} = \frac{1}{C}\frac{1}{N}\sum_{i}^{C}\sum_{j}^{N}\left|\varepsilon_{PE_{ij}}\right|\frac{L_j + V_j}{l_{ij} + v_{ij}} \qquad \text{Eq IV.A.4-7b}$$

Top/Bottom Specification (Equation varies):

$$\left| \frac{\text{Residual}}{\text{Specification Value}} \right| = \varepsilon_{Top/Btm} \qquad \text{Eq IV.A.4-8}$$

Note: If the top/bottom specification is heat input/removal, this is a special case: If Q=0 is specified, then Eq IV.A.4-8 would cause division by zero. If Q=constant (≠0) is specified, then the large value of the energy residual causes ε to be orders of magnitude too high. So for this case, the equation is changed to: | residual | $\infty 10^{-4}/V$.

IV.B. Extension to Design Reliability

This section extends the distillation model to design reliability. First, the inequality constraints are discussed. There are two types: (1) Separation requirements, and (2) Operating restrictions. Second, the process uncertainties are discussed. There are three types: (1) External, (2) Operating, and (3) Database. Finally, a convenient way of handling control variables is explained.

IV.B.1. Inequality Constraints

This section discusses inequality constraints for the distillation unit operation. There are two general types: (1) Separation requirements, such as product purities, and (2) Operating limits, such as column flooding. Generally, these limits work in opposition—maximum separation is obtained at the maximum operating limits.

In addition to the model equality constraints shown in Table IV.A.2.a-2, the design problem has additional specifications to meet. For instance, the product streams might have to meet minimum purity or component recovery requirements. Or the overhead vapor condensing temperature must be above some minimum so it can be condensed with cooling water. Likewise, the bottoms must be below a maximum temperature so condensing steam can provide the heat.

There are other physical limitations. The condenser and reboiler can only exchange limited amounts of heat. And reflux and product pumps have maximum flowrate/head relationships. Likewise, all equipment must operate at or below their design temperatures and pressures for safety. Of course, component flowrates cannot be less than zero.

Flooding occurs in the column when the vapor and liquid flows reach a certain maximum. So maximum flows are limited by the column internals as well as by heat exchanger limitations.

The preceding are all examples of “hard” constraints, i.e., constraints that must be satisfied no matter what. There are also “soft” constraints, such as minimum temperature approaches in a heat exchanger. Violating these soft constraints might be poor design, but is acceptable in practice if required to meet the hard constraints. This research makes no distinction between the two.

All these requirements can be written as inequality constraints which are, by convention, less than or equal to zero when satisfied ($g_i \leq 0$), and greater than zero when violated ($g_i > 0$).

There are two types of inequality constraints:

(1) Separation Requirements (Product Specifications)

2) Operating Restrictions

In general, these two types are in opposition and intimately connected to the operating rate (vapor and liquid internal flowrates). The separation constraints set a minimum operating rate, and the flooding constraint sets a maximum. The optimum operation satisfies the separation requirements exactly for the minimum energy usage. But for design reliability estimation we are only concerned whether or not the separation can be met. Therefore, can the column meet the design specifications while operating at its maximum rate? This latter viewpoint is useful for handling the control variables, see section IV.B.3. However, first some common types of inequality constraints are listed.

IV.B.1.a. Separation Requirements (Product Stream Specifications)

Since the whole point of distillation is to separate a chemical mixture, the separation specifications are the most fundamental design constraints. Some basic types are:

1) Minimum purity or maximum impurity of a particular component in the overhead or bottoms product stream

2) Recovery or split of a particular component

3) Ratio of two components in a product stream

4) Minimum or maximum flowrate of a particular component in a product stream

5) Minimum or maximum flowrate of a product stream

These specifications can be written simply in the form $g_i \leq 0$. For example, the minimum overhead purity for component i can be written:

$$\frac{l_{i,D}}{D} \geq Specified\,Mole\,Fraction \qquad \text{Eq IV.B.1.a-1a}$$

or

$$g_i \equiv Specified\,Mole\,Fraction - \frac{l_{i,D}}{D} \leq 0 \qquad \text{Eq IV.B.1.a-1b}$$

This demonstrates the mathematical formulation. Any product specification can be formed by writing such an inequality constraint equation.

IV.B.1.b. Operating Restrictions

Some basic forms of operating restrictions include:

1) Column operating at or below its flooding point
2) Maximum heat addition/removal from the external heat exchangers
3) Maximum pumping flowrates and head for the external pumps
4) Maximum operating temperature and pressure for the column and all peripheral equipment

Flooding is discussed in section IV.B.2.e. For this research, it is specified as a simple maximum vapor flowrate at the bottom stage, i.e.:

$$V_n \leq Maximum\,Flowrate \qquad \text{Eq IV.B.1.b-1a}$$

or

$$g_i \equiv V_n - Maximum\,Flowrate \leq 0 \qquad \text{Eq IV.B.1.b-1b}$$

Maximum Heat Transfer to or from a heat exchanger is a complicated phenomenon, dependent on many factors. For this research, a simple maximum value is specified, e.g.:

$$g_i \equiv Q_n - Maximum\,Heat\,Transfer \leq 0 \qquad \text{Eq IV.B.1.b-2}$$

IV.B.2. Process Uncertainties

This section describes various sources of distillation process uncertainties. In this research, only feed composition, tray efficiency, and database uncertainty are considered.

Distillation uncertainties can be divided into three types: (1) External, (2) Operating, and (3) Database. The significant external uncertainty is the feed composition. Feed stream state (temperature, pressure and enthalpy) is usually less significant. The two major operating uncertainties are flood point and tray efficiency. Only the latter is included in this research.

The database uncertainties are K-value ($\mathbf{K}$) and enthalpy ($\mathbf{H}$). Other database uncertainties include vapor and liquid densities, viscosities, and thermal conductivities, but their only importance is their effect on the operating uncertainties (tray efficiency and flood point). And these relationships are poorly modeled. In this research, only K-value uncertainty is included.

Finally, the external equipment, such as pumps and heat exchangers, have their own uncertainties. This is discussed in section IV.B.2.f., however this is secondary to the main problem and not included in this research.

IV.B.2.a. Feed Stream Uncertainty

The feed stream(s) composition and thermodynamic state is likely to be a major uncertainty. For testing purposes, thermodynamic state uncertainty is ignored. Feed stream uncertainty can be handled by C component flowrates, or C-1 mole fractions plus total flowrate. The former is used.

Feed stream composition uncertainty can be handled in one of two ways: C component flowrates, or C-1 mole fractions plus total flowrate. One problem with the latter is the arbitrary selection of which component mole fraction is dependent on the other C-1 mole fractions, since they sum to one. Also, there are scaling difficulties, as the C-1 mole fraction parameters vary from zero to one, while the Cth parameter for total flowrate has only a lower bound of zero. On the other hand, component flows all have the same lower bound of zero and no upper bound.

For these reasons, C uncertain parameters were used to describe the C component flowrates. For simplicity, independent, normal distributions were used for each, i.e.:

$$f_{ij} = \mathbf{N}\left\{\bar{f}_{ij}, \sigma^2_{f_{ij}}\right\} \qquad \text{Eq IV.B.2.a-1}$$

where

f_{ij} is the feed flowrate of component i to distillation stage j

$\mathbf{N}\{\cdot,\cdot\}$ is the normal probability distribution

$\bar{f}_{ij}$ is the mean value of the feed flowrate of component i to distillation stage j

$\sigma^2_{f_{ij}}$ is the variance of the feed flowrate of component i to distillation stage j

This representation is satisfactory for testing purposes, but is not intended to be accurate. For example, since the normal distribution is unbounded, it could theoretically generate a nonzero flowrate, which is impossible. The problem was avoided by specifying means and variances sufficiently far from zero.

IV.B.2.b. Tray Efficiency Uncertainty

Tray efficiency is a critical uncertainty in distillation design. For this research, a simple uniform probability distribution is used.

Tray efficiency is difficult to estimate, and there are known problems with the mathematical models. Consequently, uncertainties may be very high. And tray efficiency is likely to be a critical parameter. For this research, tray efficiency was assumed to have a uniform probability distribution, with known minimum and maximum values, i.e.:

$$\eta = \mathbf{U}\left\{\eta^{Min}, \eta^{Max}\right\} \qquad \text{Eq IV.B.2.b-1}$$

where

η — Tray efficiency

$\mathbf{U}\{\cdot,\cdot\}$ — uniform distribution

η^{Min} — Lower bound tray efficiency

η^{Max} — Upper bound tray efficiency

IV.B.2.c. Thermodynamic Database Uncertainty

Only K-values are treated as uncertain in this research. All uncertainties are handled in the Modified Wilson liquid solution model. This requires C(C-1) uncertain parameters.

As can be seen from Table IV.A.2.a-2, the two thermodynamic values which affect the process model equations are enthalpy ($\mathbf{H}$) and K-value ($\mathbf{K}$). In turn, these values are estimated from complex, empirical equations of state (EOS), which are fitted to available data. Consequently, enthalpies and K-values are uncertain. K-values are critical to the separation of the components. Enthalpies are important for external heat duties and internal column profiles and are ignored for testing purposes.

K-value is a function of temperature, pressure, and composition, i.e.:

$$\mathrm{K}_{ij} = \mathrm{K}\left(T_j, P_j, \boldsymbol{x}_{ij}, \boldsymbol{y}_{ij}\right) \qquad \text{Eq IV.B.2.c-1}$$

where subscript i denotes the i'th component, j denotes the j'th stage, and $\boldsymbol{x}$ and $\boldsymbol{y}$ are the mole fractions:

$$\boldsymbol{x}_{ij} = \frac{\mathrm{l}_{ij}}{\sum_{i=1}^{C} \mathrm{l}_{ij}} \qquad \boldsymbol{y}_{ij} = \frac{\mathrm{v}_{ij}}{\sum_{i=1}^{C} \mathrm{v}_{ij}} \qquad \text{Eq IV.B.2.c-2}$$

From thermodynamic analysis, K-value may be separated as follows (dropping subscript j):

$$\mathrm{K}_i = \frac{\gamma_i \phi_i^s P_i^{sat} \mathrm{F}_{P_i}}{\hat{\phi}_i P} \qquad \text{Eq IV.B.2.c-3}$$

where

γ_i — Activity coefficient for liquid phase from liquid solution model

ϕ_i^s — Fugacity coefficient of pure component i at T and P_i^{sat} from vapor solution model

P_i^{sat} — Pure component vapor pressure

F_{P_i} — Poynting correction factor

$\hat{\phi}_i$ — Fugacity coefficient from EOS

P — Pressure

Each part of Eq IV.B.2.c-3 is estimated from specialized equations. A rigorous treatment would analyze each part individually for uncertainty. However, this would require a vast number of uncertain parameters, which is impractical. A more reasonable approach is to realize that of the three primary sources of uncertainty— component vapor pressures, vapor phase nonidealities, and liquid phase nonidealities—the latter is generally the most significant. Therefore, it is a good approximation to assume that the only source of uncertainty is the liquid phase, and to handle all uncertainties within the liquid phase solution model parameters.

Note that this approach does not erroneously ignore any uncertainties and thermodynamic model errors. It simply provides for them only within the liquid phase solution model.

For this research, all testing was done with atmospheric distillation columns. Therefore, the vapor phase is practically ideal and for simplicity the K-value equation may be simplified to:

$$\mathrm{K_i} = \frac{\gamma_\mathrm{i} P_\mathrm{i}^{\mathrm{sat}}}{P} \qquad \text{Eq IV.B.2.c-4}$$

Note that this simplification is solely for computational speed and convenience. It is not a necessary part of the uncertainty handling and would not be done at higher pressures.

The component vapor pressure is estimated by the Miller equation (Miller, 1964):

$$ln\left(P_\mathrm{i}^{\mathrm{sat}}\right) = \frac{\mathrm{A_i}}{T} + \mathrm{B_i} + \mathrm{C_i}T + \mathrm{D_i}T^2 \qquad \text{Eq IV.B.2.c-5}$$

where T is the temperature. See Table C-1 for coefficient values for each component.

The Modified Wilson solution model (Tsuboka and Katayama, 1975) was used for the liquid phase γ_i. The correlation is:

$$ln(\gamma_i) = -ln\sum_{j=1}^{C} \mathbf{x}_j \Lambda_{ji} - \sum_{k=1}^{C} \frac{\mathbf{x}_k \Lambda_{ik}}{\sum_{j=1}^{C} \mathbf{x}_j \Lambda_{jk}} + ln\sum_{j=1}^{C} \mathbf{x}_j \frac{V_j}{V_i} + \sum_{k=1}^{C} \mathbf{x}_k \frac{V_i/V_k}{\sum_{j=1}^{C} \mathbf{x}_j \frac{V_j}{V_k}} \qquad \text{Eq IV.B.2.c-6}$$

where

$$\Lambda_{ij} = \frac{V_i}{V_j} exp\left(\frac{-\lambda_{ij} - \lambda_{jj}}{RT}\right) \qquad \text{Eq IV.B.2.c-7}$$

Λ_{ij} is an intermediate calculation.

R is the ideal gas constant.

λ_{ij} are the solution model coefficients or parameters, which are to be treated as uncertain.

V_i is the liquid molar volume of component i, calculated by the Hankinson-Thomson correlation (Hankinson and Thomson, 1979):

$$V_i = V_R^{(0)}\left(1-\omega V_R^{(\delta)}\right)V^* \qquad \text{Eq IV.B.2.c-8}$$

$$V_R^{(0)} = 1+a(1-T_R)^{1/3}+b(1-T_R)^{2/3}+c(1-T_R)+d(1-T_R)^{4/3} \qquad 0.25 < T_R < 0.95 \qquad \text{Eq IV.B.2.c-9}$$

$$V_R^{(\delta)} = \left(e+fT_R+gT_R^2+hT_R^3\right)\Big/\left(T_R-1.00001\right) \qquad 0.25 < T_R < 1.0 \qquad \text{Eq IV.B.2.c-10}$$

a	-1.52816	e	-0.296123
b	1.43907	f	0.386914
c	-0.81446	g	-0.0427258
d	0.190454	h	-0.0480645

T_R is the reduced temperature T/T_C, where T_C is the critical temperature

V^* is the component characteristic volume. See Table C-3 for component values.

ω is the acentric factor

Component molar volumes and critical temperatures are precisely known in comparison to the solution model nonidealities. Therefore, the K-value uncertainty may be handled solely within the liquid solution model coefficients (λ). Each component pair has two solution model coefficients: λ_{ij} and λ_{ji} for components i and j. Each component pair is independent of all other pairs, and thus may be analyzed separately. For C components, there are $C(C-1)$ parameters.

The thermodynamic vapor-liquid equilibria data are analyzed by nonlinear regression to estimate a confidence interval for λ_{ij}. For this research, λ_{ij} is described by a multivariate normal distribution. The SYNDES simulator (SYNDES, 1987) is available for this calculation.

Each component pair (λ_{ij} and λ_{ji}) is very highly correlated. Correlation coefficients are usually on the order of -0.95 to -0.99. Thus, in the 2-dimensional space for each

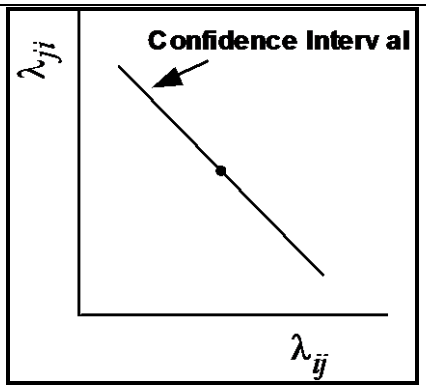


Figure IV.B.2.c-1
Modified Wilson Solution Parameters are Highly Correlated

component pair, the only statistically significant region is a 1-dimensional straight line, see Figure IV.B.2.c-1.[1]

IV.B.2.d. Column Pressure

Column pressure uncertainty is not treated in this research. However, it may easily be included using two uncertain parameters: (1) Condenser pressure drop, and (2) pressure drop per stage.

Uncertain pressure will not be developed in this research. However, it can be included later with little difficulty. This section suggests a simple approach.

Pressure must be known at each stage for the distillation simulation, and for this the condenser pressure drop and column pressure drop per stage must be estimated. Condenser pressure drop can be estimated by experience. Pressure drop per stage is difficult to estimate, but approximate, empirical correlations are available (Robbins, 1991; Kister and Gill, 1991). Pressure drop is affected by vapor and liquid internal flow, tray or packing type, and fluid physical properties.

If pressure is included as an uncertainty, the simplest approach is to use only two uncertain parameters: (1) condenser pressure drop, and (2) stage pressure drop. Including uncertainties for all parts of the pressure drop correlations (e.g., density, viscosity, etc.) would require a large number of parameters.

IV.B.2.e. Column Flood Point

Column flood point was not included as an uncertainty in this research. However, flood point is the most critical 'internal' uncertainty, along with tray (or packing) efficiency, as it sets the maximum allowable operation for the column. Present correlations are empirical and underdeveloped. For simplicity, flood point was set at some fixed, maximum vapor reboil rate.

The distillation column has internal operating limitations because of the hydraulics of the tray or packing, see Figure IV.B.2.e-1. Too low a vapor flowrate and the column dumps or weeps. Too high a vapor flowrate and the column blows or floods. Likewise, the liquid has minimum and maximum flow restrictions. To complicate matters, different tray and packing types have different regions of operability. Also, tray efficiency and pressure drop vary throughout the region of operability.

---

[1] The confidence interval is actually an ellipse. However, when drawn to scale, the minor principal axis is practically zero.

For design reliability estimation we are only interested in maximum flowrates, because this is where the column operates most effectively (maximum separation). A column fails only if it cannot perform a separation at its maximum operating rate.

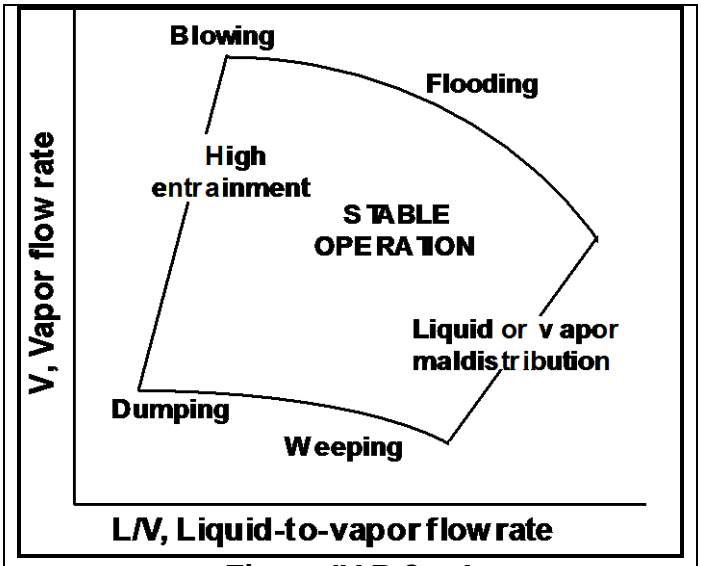


Figure IV.B.2.e-1
Tray Malfunctions as a Function of Loading (Henley and Seader, 1981, p54)

Despite the vast experience with distillation, it is difficult to predict the flooding point of a column accurately. Empirical methods exist (Henley and Seader, 1981, Ch 13; Dankworth and Sundaresan, 1989; Kister and Gill, 1991; Robbins, 1991; Rocha et al., 1991; Bolles and Fair, 1979), but these techniques are highly uncertain and generally conservative. A rigorous statistical examination has yet to be undertaken.

For simplicity, this research assumes that flooding occurs at a single maximum vapor flowrate at the bottom stage with no uncertainty. More accurate models that allow for uncertainty may be added later.

IV.B.2.f. External Equipment (Heat Exchangers, Pumps)

This section discusses uncertainties associated with the external equipment, their effect on distillation column reliability, and how they may be handled. However, these uncertainties are not included in this research.

Each distillation column has external equipment, including reboiler and condenser heat exchangers and feed, overhead, reflux and bottoms pumps, see Figure IV.A.1-1. This section considers the additional uncertainties caused by the external equipment, and how they affect the distillation column operation.

There are two approaches for handling the external equipment:

Simultaneous — Considering the column and external equipment as a single operation.

Separated — Analyzing each piece of equipment separately.

The simultaneous approach is straightforward, but increases the size of the problem as additional parameters and inequality constraints are included in the analysis. The separated

approach produces smaller subproblems, but requires careful handling of the interrelationships.

Pumps

Pumps are well developed process operations and therefore should not introduce additional uncertain parameters. However, new inequality constraints are added that specify a maximum flowrate-head relationship for the external streams.

Since pump safety factors are economically trivial in comparison to distillation column safety factors, it may be easiest to design the pumps to handle the maximum operating conditions, thus guaranteeing that the pumps will not limit distillation column operation. This may also be handled statistically, e.g., by calculating the 95% or 99% 'worst-case' conditions after the distillation column design is finalized.

Heat Exchangers

The fundamental uncertain parameter for heat exchanger design is the heat transfer coefficient. This must be estimated by empirical correlations and previous experience, however this is fairly well developed in comparison to the distillation uncertainties.

Heat exchanger design sets the maximum heat transfer to and from the reboiler and condenser, thereby restricting the maximum internal vapor and liquid flowrates of the distillation column. This may be less than or greater than the column flood point depending on the design specifics. If this is greater than the flooding point, then the heat exchanger contributes no additional uncertainty to the column design and may be ignored. However, this may not always be economically appropriate because heat exchangers are expensive.

Using the simultaneous approach, heat exchangers may be included in the overall reliability analysis by treating the reboiler and condenser heat transfer coefficients as uncertain parameters, and adding inequality constraints limiting the maximum heat transfer in the reboiler and condenser. This is the most reasonable approach. The separated approach is trickier, because it depends on whether or not the heat exchanger design is limiting or not, and the reliabilities of the column and heat exchanger equipment are interrelated.

IV.B.3. Elimination of Control Variables (Degrees of Freedom) by Constraint Substitution

Distillation control variables are handled by turning two of the inequality constraint specifications into equality constraints, used for the process model simulation specifications. Pressure is set at atmospheric conditions. This avoids the difficulty of optimizing control variables in the line-search algorithm.

While control variables (z) are degrees of freedom that provide operating flexibility, they complicate the boundary point line-search procedure. Luckily, there is an automatic way to optimize them without direct manipulation within the line-search procedure. As proved by Grossmann and Floudas (1987), $q+1$ inequality constraints are always active, where q is the number of control variables. If we can determine *which* inequality constraints are active, these may be set to equalities and used as problem specifications instead to solve the simulation model (section IV.A.2.c).

For distillation, control variables include column pressure plus two others. However, separation usually improves as pressure decreases (except for certain nonideal thermodynamic systems, and assuming temperature remains within condenser and reboiler heat transfer restrictions), so we can usually set column pressure to atmospheric conditions and remove it as a control variable.[1] This leaves two control variables to be substituted by active inequality constraints.

The most obvious choice is the maximum operating capacity, since this maximizes separation and is always active. Therefore, one of the distillation problem specifications should be whichever of the following is limiting:

$$Q_l = Q_l^{Max} \quad (\text{specified quantity}) \qquad \text{Eq IV.B.3-1a}$$

or

$$Q_n = Q_n^{Max} \quad (\text{specified quantity}) \qquad \text{Eq IV.B.3-1b}$$

or

$$V_n = V^{Flood} \quad (\text{specified quantity})^{2} \qquad \text{Eq IV.B.3-1c}$$

These specifications are simplified to be independent of the uncertain parameters θ. Therefore, they may be set and ignored during the boundary point line-search. When more complicated correlations are used for the heat transfer rates and flood point (functions of θ), an extra step is required: For each value of θ in the line-search iteration, first calculate the flood point and maximum heat transfer rates, then use the limiting value for the distillation problem specification.

Any one of the remaining inequality constraints may be used for the second active constraint specification, so long as it is not always inactive (i.e. always 'behind' another

---

1 Technically, minimum pressure at atmospheric conditions becomes an active inequality constraint.

2 Assuming that flooding occurs at the bottom stage.

inequality constraint). An obvious choice is to use one of the key component separation specifications. Different choices will result in different boundary geometries, although this aspect was not studied. Note: If a inactive inequality constraint is chosen by mistake, simulation will quickly show that the true active inequality constraint is violated.

IV.B.4. Finite-Differencing for Gradient Approximation

This section discusses the step-sizes used for the finite-difference gradient approximation at the converged boundary point for distillation.

The nested-loop algorithm (section III.E) does not automatically provide the gradient information required for the boundary-approximation (unlike the simultaneous algorithm). Therefore, these gradients must be calculated manually. Because distillation is solved iteratively, finite-differencing is required which unfortunately is inexact: Too small a step, and computer numerical roundoff occurs. Too large a step, and the first-order approximation is inaccurate. Accuracy is also affected by the distillation convergence tolerance (section IV.A.4).

Computer numerical roundoff is negligible in comparison to iterative solution error. And second-order curvature should be insignificant for the step sizes used in practice. Therefore, distillation convergence tolerance is the major factor. For this research, highest accuracy convergence tolerances were used, since accuracy was paramount. However, there is much room for future optimization. For example, accuracy may be maintained while reducing computational effort by increasing the step size and loosening the convergence tolerance accordingly.

The finite-difference step sizes used for this research are as follows:

Feed Flowrate Parameters

Feed flowrates may vary from zero to very large numbers. Step distances were originally specified as a percentage of the component feedflow rate, e.g., 0.1% of component flow. However, this does not work for trace components—if the total flowrate is say 10,000 lbm/hr, varying a trace component from 0.1 lbm/hr to 0.1001 lbm/hr will probably have insignificant effect and incorrectly result in gradients of zero. Instead the step distance for all components was set at 0.1% of the feed stream *total* flowrate.

Tray Efficiency Parameters

Simulation studies showed that second-order curvature was significant at steps greater than 1% tray efficiency. +0.1% step distance was used.

Thermodynamic Database Parameters

Modified Wilson parameters (Tsuboka and Katayama, 1975) may be positive or negative, zero or large. A percentage-type step distance fails when the parameter base value is zero. However, a percentage-type step is appropriate for large values. 0.1% of base value or 0.01 was used, whichever is larger.

For all test cases, the step distances are sufficiently large for computer roundoff to be negligible. Second-order curvature is negligible. Roundoff error due to iterative convergence is by far the most important factor. Gradient calculations are estimated to be within 1-2% accuracy at worst. One special programming modification is noteworthy: If the gradient is small for a particular parameter, the distillation simulation occasionally converges in the initial iteration or after only one Newton step. This can introduce gradient error. For safety, the distillation simulator is forced to perform at least two Newton steps in the finite-difference calculation.

V. Testing of the Boundary-Approximation Procedure

V.A. Geometric Test Problems: Preliminary Development of the Boundary-Approximation Procedure

This section discusses the earliest studies, when the procedure was tested using simple geometric test problems. The purpose was to demonstrate that the procedure converges accurately. This was successfully done.

The boundary-approximation procedure was first tested using 2-dimensional geometric test problems. It was found that the reliability converged quickly to the correct value.

Next, a 3-dimensional test problem was developed. However, it was found that the optimistic and in particular the pessimistic reliability estimates converged very slowly and asymptotically. Additional geometric test problems would be of little value and 'real' test problems were required with realistic constraint boundaries. No further geometric problems were tested.

All the geometric test problems used the simultaneous line-search algorithm described in section III.E.2 to obtain the boundary points.

This section only tests the optimistic and pessimistic boundary-approximations as the best-estimate boundary-approximation had yet to be conceived of.

V.A.1. Simple 2-D Test Problem

The first test problem was in 2-dimensions. Four boundary equations were defined. All were convex. For simplicity, the equations are functions of only θ (not x or z). Table V.A.1-1 lists the equations. Figure V.A.1-1 plots the constraint boundary. There are 4 constraints (g_1 to g_4). g_1 and g_2 are symmetrical, but the design point θ^{Des} was moved off the origin to make it unsymmetrical. Also, g_4 is totally inactive which is a possible situation.

The design point (θ^{Des}) is (0.4, 0.1). The axial boundary points were obtained as the starting points for initialization. In Figure V.A.1-1, these are points 1,2,3,4. Next, the orthogonal points were obtained, shown as points 5,6,7,8. The boundary points are also listed in Tables V.A.1-2 and V.A.1-3.

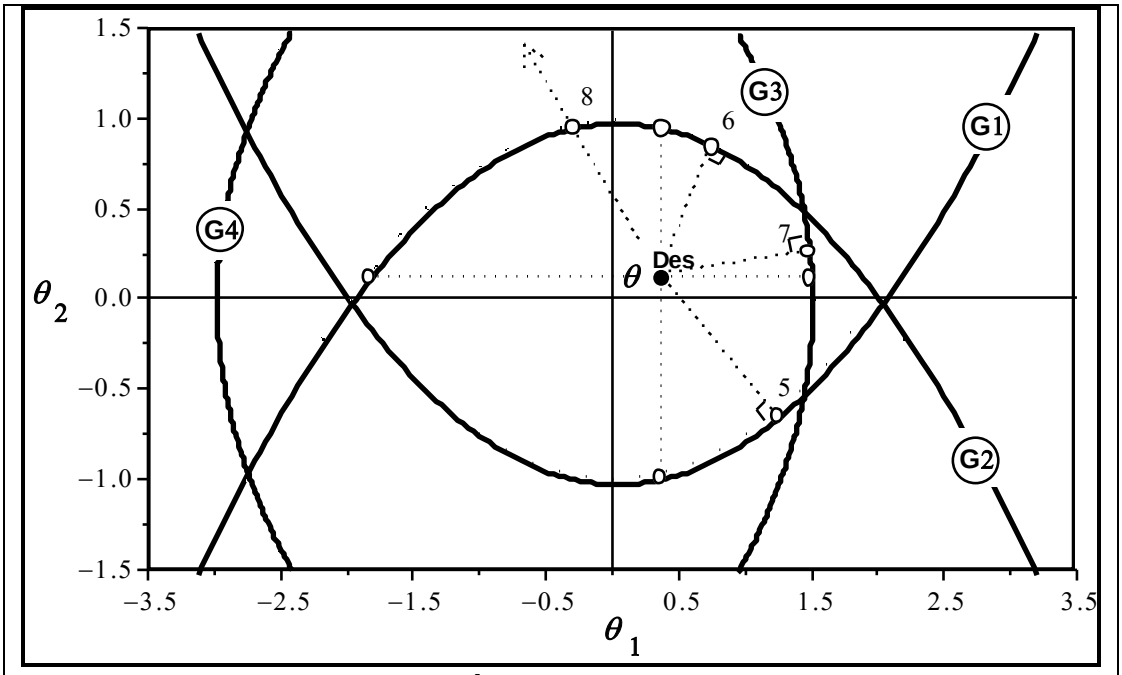


Figure V.A.1-1
Simple 2-D Test Problem Constraint Boundary

Table V.A.1-1
Simple 2-D Test Problem: Constraint Equations

Constraint	Equation
G1	$\frac{1}{4}\theta_1^2 - \theta_2 - 1$
G2	$\frac{1}{4}\theta_1^2 + \theta_2 - 1$
G3	$\theta_1 + \frac{1}{4}\theta_1^2 - 1.5$
G4	$-\theta_1 + \frac{1}{4}\theta_1^2 - 3$

Table V.A.1-2
Simple 2-D Test Problem: Axial Direction Boundary Points

Point #	Radial Direction	Boundary Point (θ_1, θ_2)	Active Constraint
1	(+,0)	(1,4975 , 0.1)	3
2	(0,+)	(0.4 , 0.96)	2
3	(–,0)	(-1,897367 , 0.1)	2
4	(0,–)	(0.4 , -0.96)	1

Table V.A.1-3
Simple 2-D Test Problem: Orthogonal Boundary Points

Point #	Boundary Point (θ_1, θ_2)	Active Constraint
5	(0.7643129 , -0.8539565)	1
6	(0.6613624 , 0.89065)	2
7	(1,487864 , 0.2203288)	3
8	(-0.3184988 , 0.9746397)	2[1]

Orthogonality was confirmed for boundary points 5-8. For instance, g_1 at point #5 has a gradient of $(\theta_1/2, -1.0)$ or (0.3821564,-1.0). And the vector $\left(\hat{\theta}^5 - \theta^{Des}\right)$ is (0.3643129 , -0.9539565). Now these vectors are parallel, as shown by calculating the dot product or angle between vectors:

$$\text{Angle between } (\mathbf{A},\mathbf{B}) = cos^{-1}\left[\frac{\|\mathbf{A}|^{0.5}\|\mathbf{B}|^{0.5}}{\|\mathbf{AB}\|}\right] \qquad \text{Eq V.A.1-1}$$

$$= cos^{-1}\left[\frac{\left(\sum A_i^2\right)^{0.5}\left(\sum B_i^2\right)^{0.5}}{\left(\sum A_i B_i\right)}\right] \qquad \text{Eq V.A.1-2}$$

which for point #5 is:

$= cos^{-1}$ [(1.14604351)$^{0.5}$(1.04275689)$^{0.5}$ / (1.093181)]
$= cos^{-1}$ [1.000000]
= 0°

[1] Constraint 4 is inactive (behind all other constraints).

The optimistic boundary-approximation obtained from the axial and orthogonal points is shown in Figure V.A.1-2. The pessimistic boundary-approximation is shown in Figure V.A.1-3. The optimistic boundary is fairly accurate, except for a significant error in the lower left quadrant. And it completely encloses the true boundary, which guarantees that it is a true optimistic boundary. Likewise, the pessimistic boundary is rigorous.

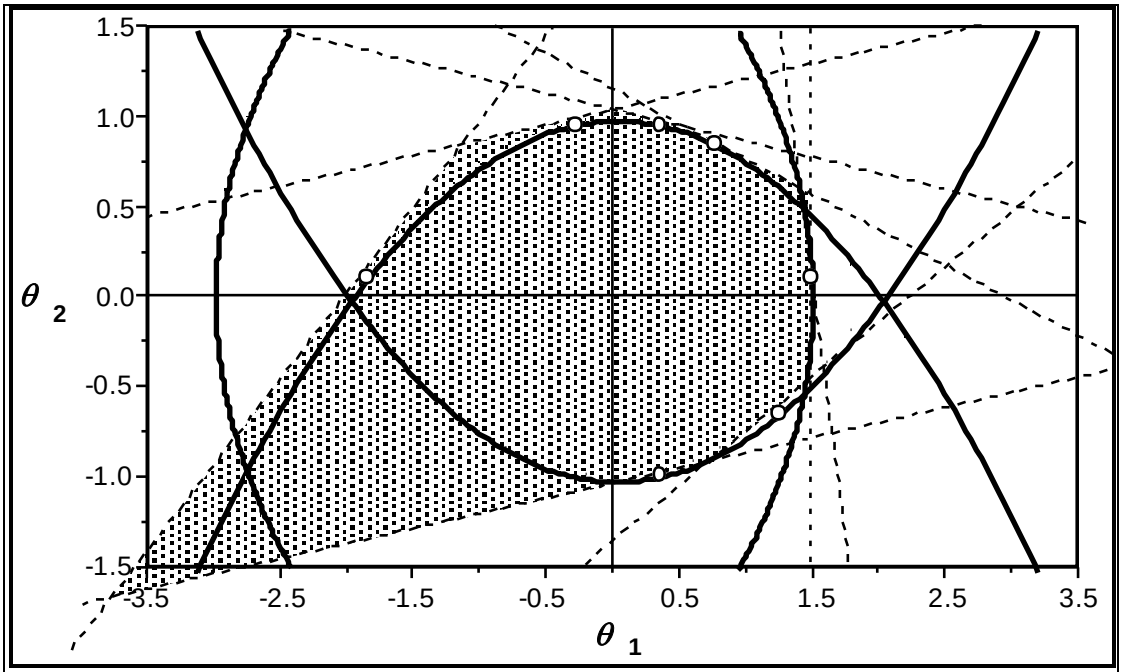


Figure V.A.1-2
Simple 2-D Test Problem: Optimistic Boundary From First 8 Points

The boundary-approximation procedure was run to 27 boundary points using 5,000 Monte Carlo integration points. Figure V.A.1-4 plots the optimistic and pessimistic design reliability estimates, along with 99% confidence intervals. The procedure converges to the true value (calculated by analytical integration), and at no time are the confidence intervals incorrect. A different sequence of Monte Carlo integration points was used for each point, which causes the fluctuation in the reliability estimates.

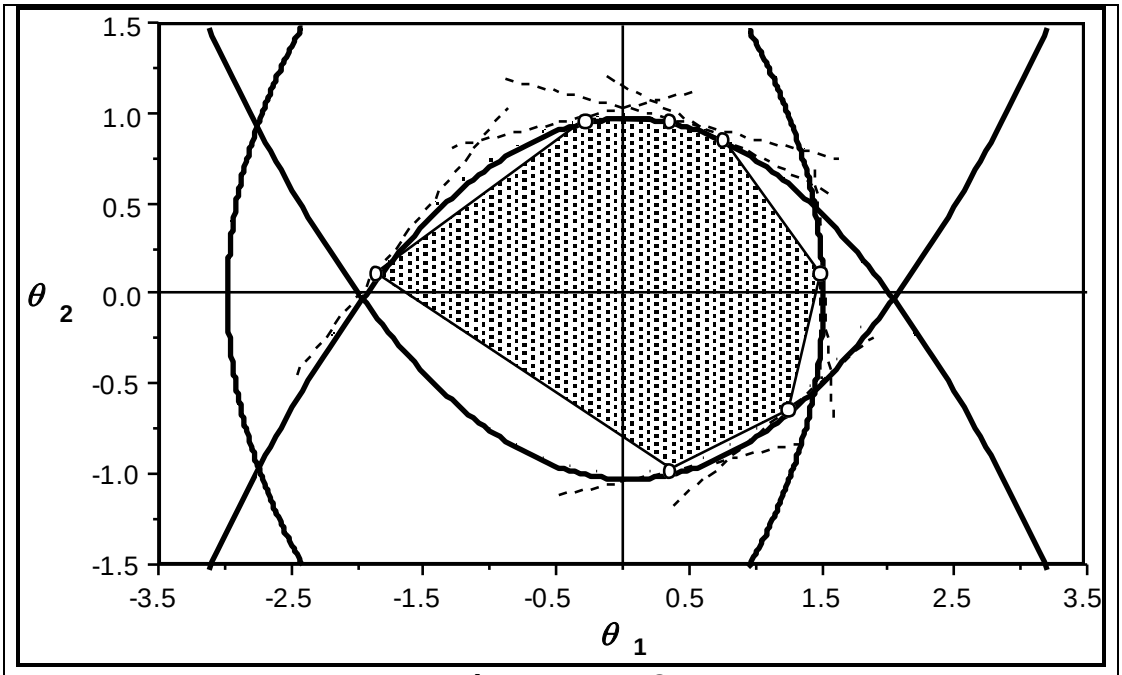


Figure V.A.1-3
Simple 2-D Test Problem: Pessimistic Boundary From First 8 Points

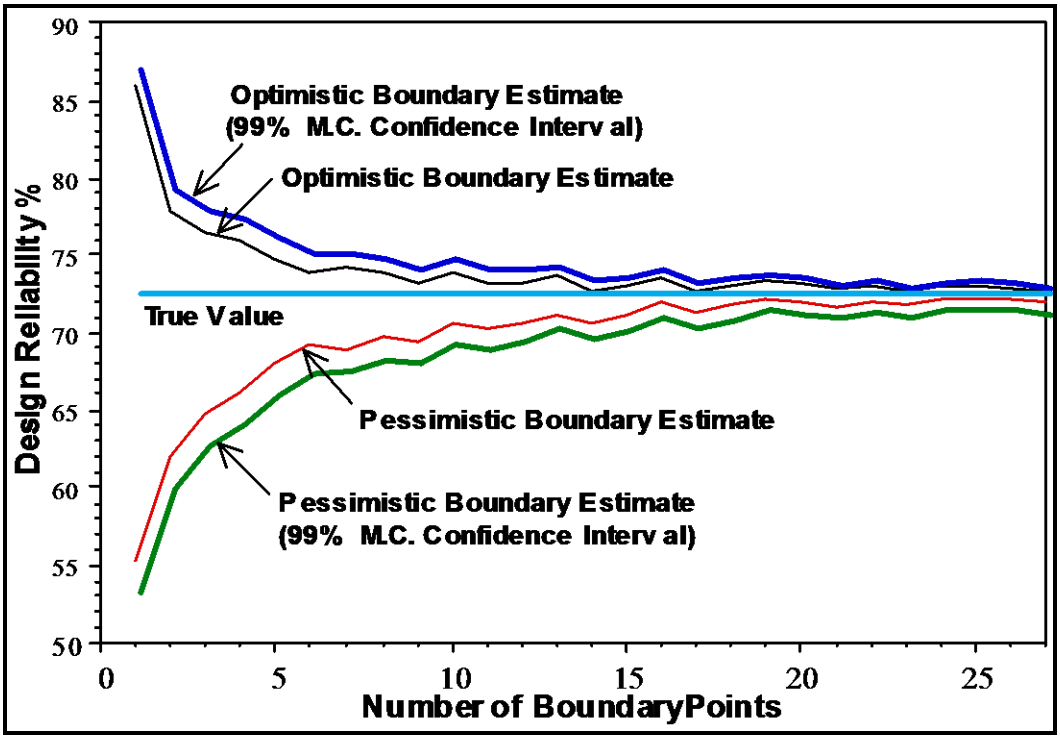


Figure V.A.1-4
Simple 2-D Test Problem Convergence

Next, the procedure was continued to 75 boundary points and performed a $4*10^6$ Monte Carlo integration, to get a precise boundary estimate. The result is shown in Figure V.A.1-5. The true value is the horizontal line, which was calculated analytically. The optimistic and pessimistic estimates bound the true value correctly. To check the code, a subroutine was written to check each point against the *actual* boundary, and performed the Monte Carlo integration. The results were accurate.

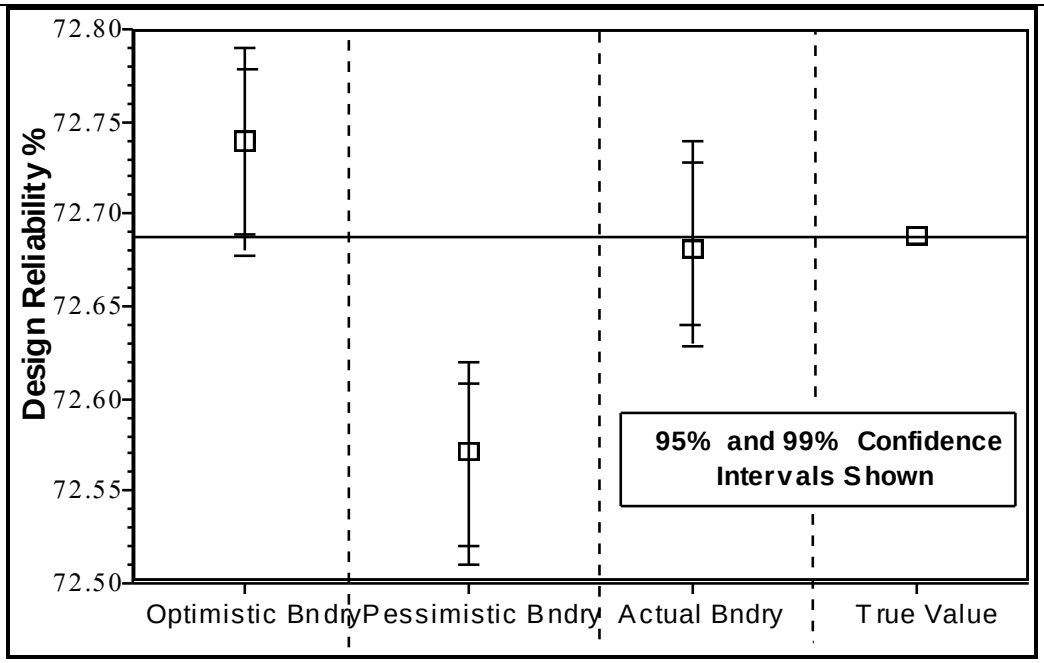


Figure V.A.1-5
Simple 2-D Test Problem: Final Estimated Design Reliability

V.A.1.a. Confirmation of the Simultaneous Algorithm When the Constraint Boundary Equation (g) is a Function of θ and $\boldsymbol{X}$

The simultaneous algorithm has been confirmed (orthogonal and radial) when the constraint boundary ($\mathbf{g}$) is a function of $\boldsymbol{\theta}$. This section confirms the simultaneous algorithm when the constraint boundary is also a function of $\boldsymbol{x}$.

The complete problem is not needed. As a starting point, g_2 was modified from the previous section. The new boundary equation is:

$$\mathrm{g} = \frac{1}{4}\theta_1^2 + \theta_2 + 1 + x_1^2 \qquad \text{Eq-V.A.1.a-1}$$

with

$$\mathrm{h}_1 = \theta_2 - x_1 = 0 \qquad \text{Eq-V.A.1.a-2}$$

Another constraint was added:

$$h_2 \;=\; \theta_1 - x_1 + x_2 \;=\; 0 \qquad \text{Eq-V.A.1.a-3}$$

which does not affect the orthogonal point, and simply calculates x_2 to satisfy h_2.

Starting with the design point of θ^{Des} = (0.4 , 0.1), the algorithm converged to an orthogonal point of $\hat{\theta}$ = (0.4508856 , 0.5950688) and $\hat{x}$ = (0.5950688 , 0.1441832). It was confirmed that $\mathbf{g}$ =0 and $\mathbf{h}$ =**0**. Orthogonality was confirmed using Eq III.E.2.a.i-17.

V.A.2. 2-D Test Problem with Concave/Convex Boundaries

To test concave/convex boundaries, the following 2-D test problem was used, Figure V.A.2-1. There are 4 constraints: g_1 and g_2 are convex. g_3 and g_4 are concave. It is shown with both axial and orthogonal boundary points.

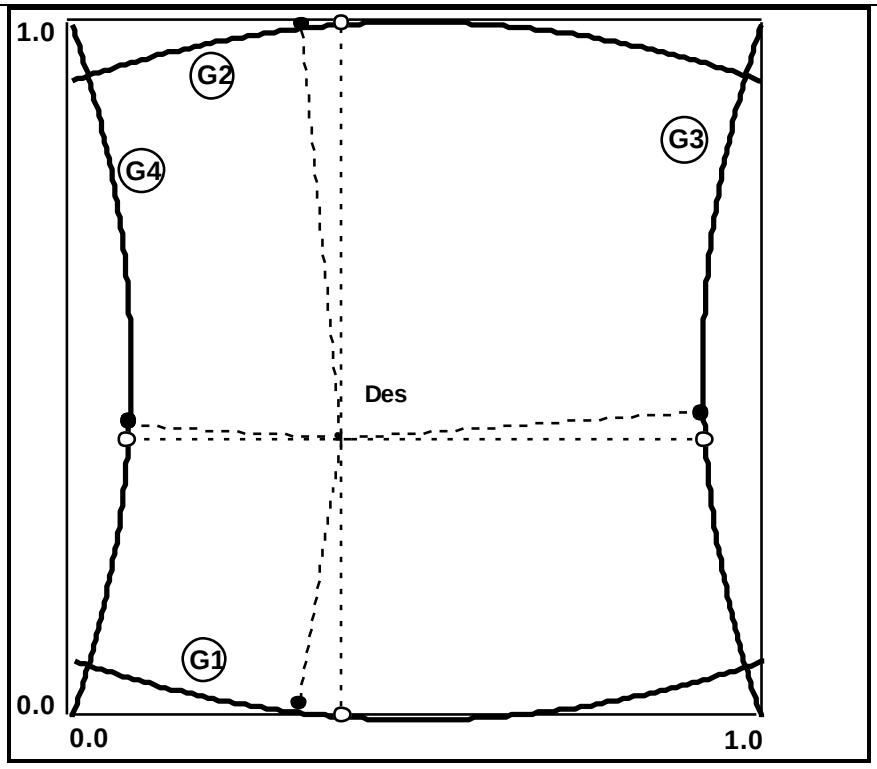


Figure V.A.2-1
Concave/Convex 2-D Test Problem Constraint Boundary

Table V.A.2-1
Concave/Convex 2-D Test Problem: Constraint Equations

Constraint	Equation
G1	$\frac{1}{5}(\theta_1 - 0.5)^2 - \theta_2$
G2	$\frac{1}{5}(\theta_1 - 0.5)^2 + \theta_2 - 1$
G3	$\theta_1 - \frac{1}{5}(\theta_2 - 0.5)^2 - 0.95$
G4	$-\theta_1 - \frac{1}{5}(\theta_2 - 0.5)^2 + 0.05$

Figure V.A.2-2 plots the optimistic and pessimistic reliability estimates as a function of the number of boundary points. The "true" value is also shown, which was calculated by performing the Monte Carlo integration with the true geometrical boundary. Note that at

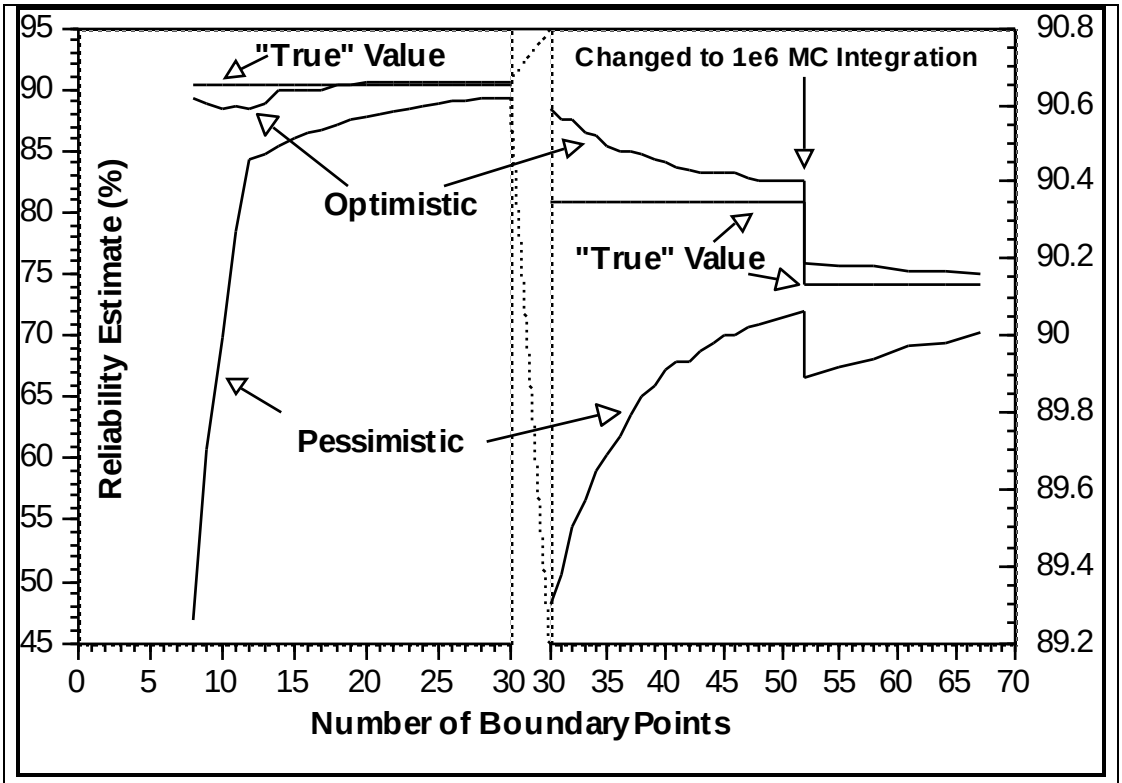


Figure V.A.2-2
Concave/Convex 2-D Test Problem Convergence

initialization, the optimistic reliability estimate is *less* than the true value. However, the optimistic and pessimistic reliability estimates converge correctly. Also note that the optimistic reliability is quite accurate, while the pessimistic reliability estimate is initially very low and converges more slowly. The initial convergence was run with 25,000 Monte Carlo integration points (same sequence repeated every additional boundary point to avoid the fluctuations in Figure V.A.1-4). At 52 boundary points, this was increased to 100,000 Monte Carlo integration points, which causes a step change. Also, the Figure is split into two y-axial ranges for better visualization.

V.A.3. 3-D Test Problem: Simple Sphere

The constraint boundary was defined as a simple sphere, with a boundary of radius 2.0, with a uniformly distributed parameter distribution in a sphere of radius 2.1. This has a 'true reliability' of 86.38%, calculated analytically.

The optimistic and pessimistic reliability estimates are plotted in Figure V.A.3-1. The pessimistic estimate starts extremely low (15%) and is very slow to converge. The region of uncertainty (optimistic-pessimistic) decreases significantly at first, but slows quickly. It will take an unacceptably large number of boundary points to converge. Convergence rates are listed in Table V.A.3-1.

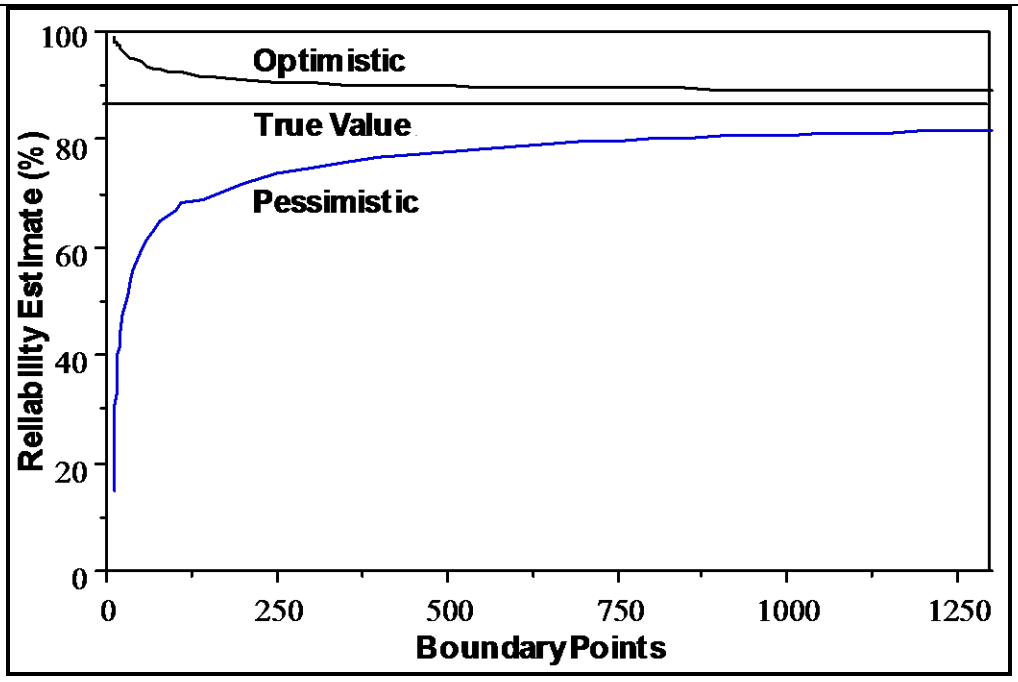


Figure V.A.3-1
3-D Spherical Test Problem: Optimistic and Pessimistic Reliability Estimates Converge Very Slowly

Table V.A.3-1
3-D Spherical Test Problem: Optimistic-Pessimistic Reliability Estimates Converge Very Slowly

# Bndry Points	Pessimistic %	Optimistic %	Opt-Pess %	% Improvement Per New Bndry Point
50	59.26	94.36	35.10	0.28
100	67.00	92.54	25.54	0.12
500	77.81	89.94	12.13	0.012
1300	81.83	88.89	7.06	0.0033

This problem is caused by the large number of 'gaps' or nearest-neighbor groups in 3-D. For instance, 500 boundary points in addition to the axial points, results in 2^3 + 500(2) = 1,008 gaps. Because of the nature of the problem (spherical boundary, with uniform distribution), each gap has a small but significant statistical uncertainty which adds up to a large overall uncertainty (e.g., 0.02% uncertainty in each gap gives 20% total uncertainty). The only way to reduce these uncertainties is to obtain boundary points for *each* gap. So we need a great many points, and each new point can only resolve a small amount of the uncertainty.

The solution was to develop the best-estimate boundary-approximation and use the optimistic and pessimistic boundary-approximations solely to quantify the uncertainty and select the next boundary point. However, while this was being developed, it was decided that the basic procedure had been demonstrated sufficiently. The next level of testing required knowledge of the actual boundary shapes for distillation problems. Consequently, geometrical testing was ended.

V.B. Distillation Test Problem Development

This section describes the general shape of the distillation constraint boundary, and develops a set of initial 'key' boundary points which are used to construct the initial boundary-approximation.

Three binary distillation operations and one ternary distillation operation were developed and tested. These were:

1) Acetone / Benzene
2) Acetone / Water
3) 2-Methyl-2-Butene-1 / Isoprene (2MB1/IPM)
4) Acetone / Benzene / Toluene

Acetone/Benzene is a fairly ideal wide-boiling system. Acetone/Water is a nonideal, very wide-boiling system. 2MB1/IPM is a very narrow-boiling system. For the ternary case, Toluene was added to the Acetone/Benzene system as an extra nonkey component to increase the dimension.

In the testing of geometric problems, the boundary-approximation initialization included the axial boundary points and the orthogonal boundary points. For distillation problems, the axial points present no problem. However, orthogonal points are difficult for two reasons:

1) The nested-loop algorithm cannot find orthogonal points without unreasonably large effort.
2) In some cases, an individual constraint can wrap completely around the focalpoint. There is no meaningful orthogonal point for this case.

Consequently, orthogonal boundary points were not used for distillation testing. The procedure was initialized using only axial boundary points and any additional points thought useful, as will be discussed.

Plots of the constraint boundaries for the distillation test problems are provided in Appendix C. Certain plots are also shown in the following sections, where appropriate.

V.B.1. Feed Component Flowrate Boundary Space

In the feed component flowrate space, the constraint boundary is 'balloon-shaped'. Each component has a maximum feed rate and may or may not have a minimum feed flowrate (besides zero).

The feed stream uncertainty is described by C parameters, one for each component flowrate. The schematic shown in Figure V.B.1-1 is a typical boundary shape.

The distillation column always satisfies the constraints at the design feed flowrate (θ^{Des}). As each component flowrate is increased, at some point the column fails. This failure could occur because a separation specification related to that particular component cannot be met, or simply because the feed flowrate becomes so large that the component 'squeezes out' the other components, and essentially floods the column. So each component has a maximum feed flowrate. This maximum may or may not be statistically significant.

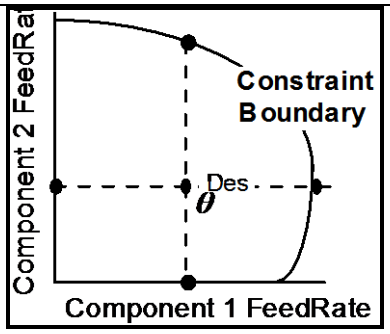


Figure V.B.1-1
General Balloon Shape of the Constraint Boundary in the Component Feed Flowrate Space

If the component is a nonkey component simply passing through the column, then its minimum feed flowrate is zero. However, it might be nonzero if connected to a process specification, such as minimum recovery or purity for that component.

For the initialization set of boundary points in the feed-space, the boundary points in the positive axial direction of each component's feed flowrate are found. In the negative direction, the zero flowrate with orthogonal gradient is used as the boundary point unless a nonzero minimum flowrate exists for that component. Figure V.B.1-2 shows a schematic of the best-estimate, optimistic and pessimistic boundary-approximations using the initialization point set.

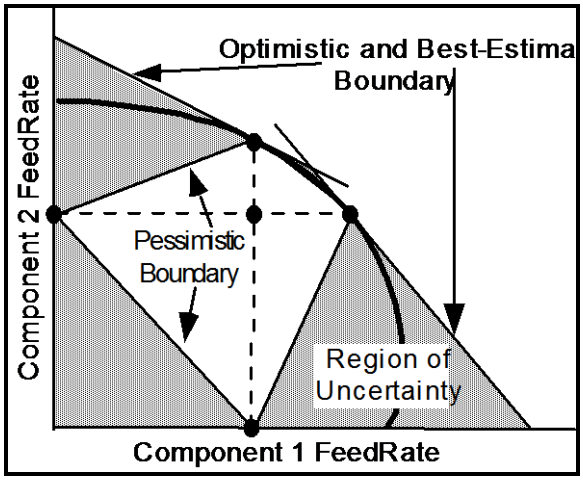


Figure V.B.1-2
Initial Boundary-Approximations in the Component Feed Flowrate Space

V.B.2. Tray Efficiency

This section discusses the general shape of the constraint boundary as a function of tray efficiency. Generally, this is a monotonic function—as tray efficiency increases, the constraint boundary is enlarged Also, since tray efficiency is treated as a uniform probability distribution with maximum and minimum values, special handling is warranted.

Separation improves as tray efficiency increases. Thus, the column can satisfy the product specification constraints while operating at lower internal flowrates. Conversely, the column can satisfy the constraints for more extreme parameter values, i.e., the constraint boundary is enlarged. This is shown schematically in Figure V.B.2-1 in the component feed space.

Note that the constraint boundary is open-ended in the positive tray efficiency direction. Typically though, 100% tray efficiency is the maximum.

This relationship is monotonic in general. However, it could be non-monotonic for multicomponent systems with unequal tray efficiencies and nonideal thermodynamics. For example, if a binary pair's relative volatility is strongly affected by a third component, then a lower tray efficiency in the third component might change the distribution sufficiently so as to improve the binary pair separation.

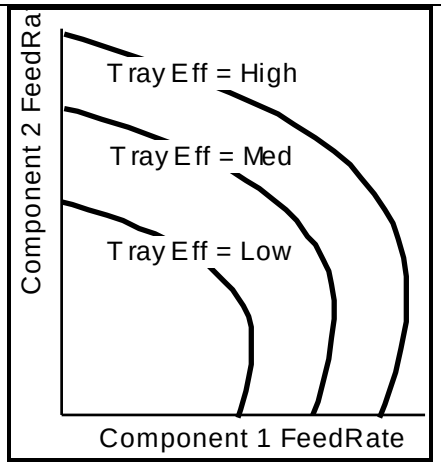


Figure V.B.2-1
Effect of Tray Efficiency on the Constraint Boundary in the Feed Flowrate Space

For this research, tray efficiency uncertainty is a uniform distribution with fixed minimum and maximum values (section IV.B.2.b.). This leads to two important questions: First, what is the best location for the focalpoint: (1) At the maximum value looking 'down', (2) in the middle, or (3) at the minimum value looking 'up'. Second, how should the boundary-approximation procedure handle min/max limits for individual parameters, since the radial coordinate system does not handle this naturally? These two questions are addressed in the following sections.

V.B.2.a. Location of Tray Efficiency Focalpoint

The focalpoint should be placed at the maximum tray efficiency value, as this improves the geometric shape of the constraint boundary and avoids hyperquadrant doubling.

There are three alternatives for choosing a tray efficiency focalpoint: (1) Maximum value, (2) Intermediate value, (3) Minimum value. This is shown schematically in Figure V.B.2.a-1.

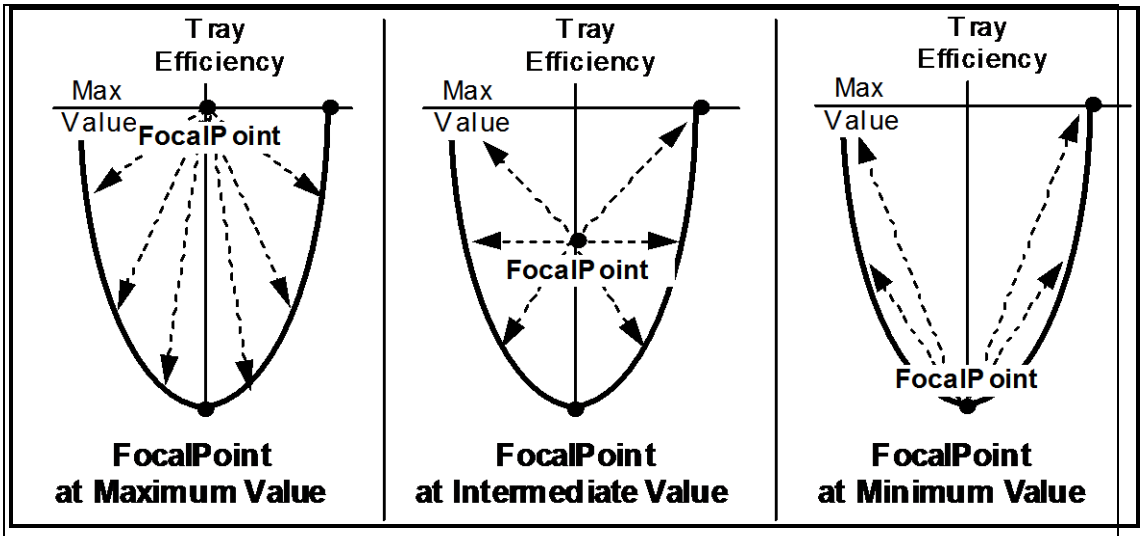


Figure V.B.2.a-1
Three Alternatives for Tray Efficiency Focalpoint: Maximum Value, Intermediate Value, or Minimum Value

The best alternative is to set the focalpoint at the maximum tray efficiency. First, the constraint boundary is closed and is more convex, which should provide a better boundary-approximation. Second, the boundary is in one direction only (downward), therefore the number of hyperquadrants—and the number of nearest-neighbor groups and computer memory—is not doubled by the additional parameter.[1]

[1] The current boundary-approximation code must be modified to take advantage of this. To save time for prototyping and testing, this was not done—the tray efficiency uncertainty parameter *does* double the computer memory. The focalpoint is set at the maximum tray efficiency value (e.g., 100%) and a dummy boundary point is specified with a slightly greater tray efficiency value (e.g., 101%). Consequently, half the neighbor groups are created in memory, but never used.

V.B.2.b. Keeping the Boundary Point Line-Search Between the Maximum and Minimum Tray Efficiency

In this work, tray efficiency has a uniform probability distribution, so there is a minimum and maximum tray efficiency. It makes sense, therefore, to constrain the boundary point search to remain between these two extremes since the rest of the boundary is of no concern. This can be handled by truncating the line-search whenever it would extend past these practical limits. This section performs a test simulation to see if convergence is improved by doing so. Results were mixed. In this research, tray efficiency *was* truncated.

Below the minimum tray efficiency, the constraint boundary is of no concern. However, the boundary-approximation procedure uses radial coordinates to select the next boundary point and may select simulation points below the minimum tray efficiency, as shown schematically in Figure V.B.2.b-1. Does this require any special handling? One option is to do nothing different—do not truncate the line-search and map the boundary-approximation over the whole region. Another option is to truncate the line-search at the minimum tray efficiency value, then check to see if the boundary is encountered at that point. If it is, continue the line-search as before. If not, define the boundary-approximation as the minimum tray efficiency horizontal plane.

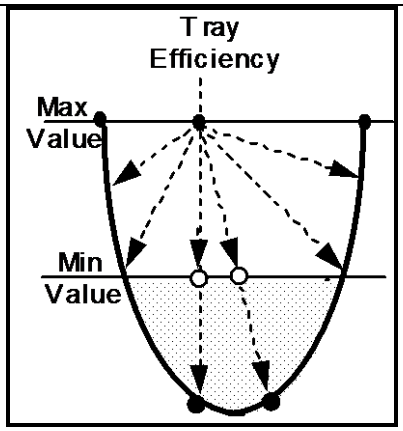


Figure V.B.2.b-1 Schematic of the Next-Point Line-Search Extending Below the Minimum Tray Efficiency

Both approaches are valid, but they have significantly different convergence characteristics. Not-truncating wastes computational effort in boundary regions of no concern, but requires fewer boundary points overall because a true boundary point, even though below the minimum tray efficiency, provides more useful boundary information than the minimum tray efficiency horizontal plane. Also, the pessimistic boundary-approximation is geometrically larger (i.e., less pessimistic), so convergence is improved. On the other hand, while truncating the line-search does not model the true boundary shape as accurately for the same number of boundary points, it could be faster because each truncated boundary point requires only one simulation instead of several.

A test was made using the acetone-benzene test problem, with feed flow and tray efficiency uncertainties ($p=3$). The procedure was converged with and without truncation. Figure V.B.2.b-2 compares the size of the uncertainty regions (optimistic-pessimistic) and converging reliability estimate error as a function of the number of boundary points and execution time.

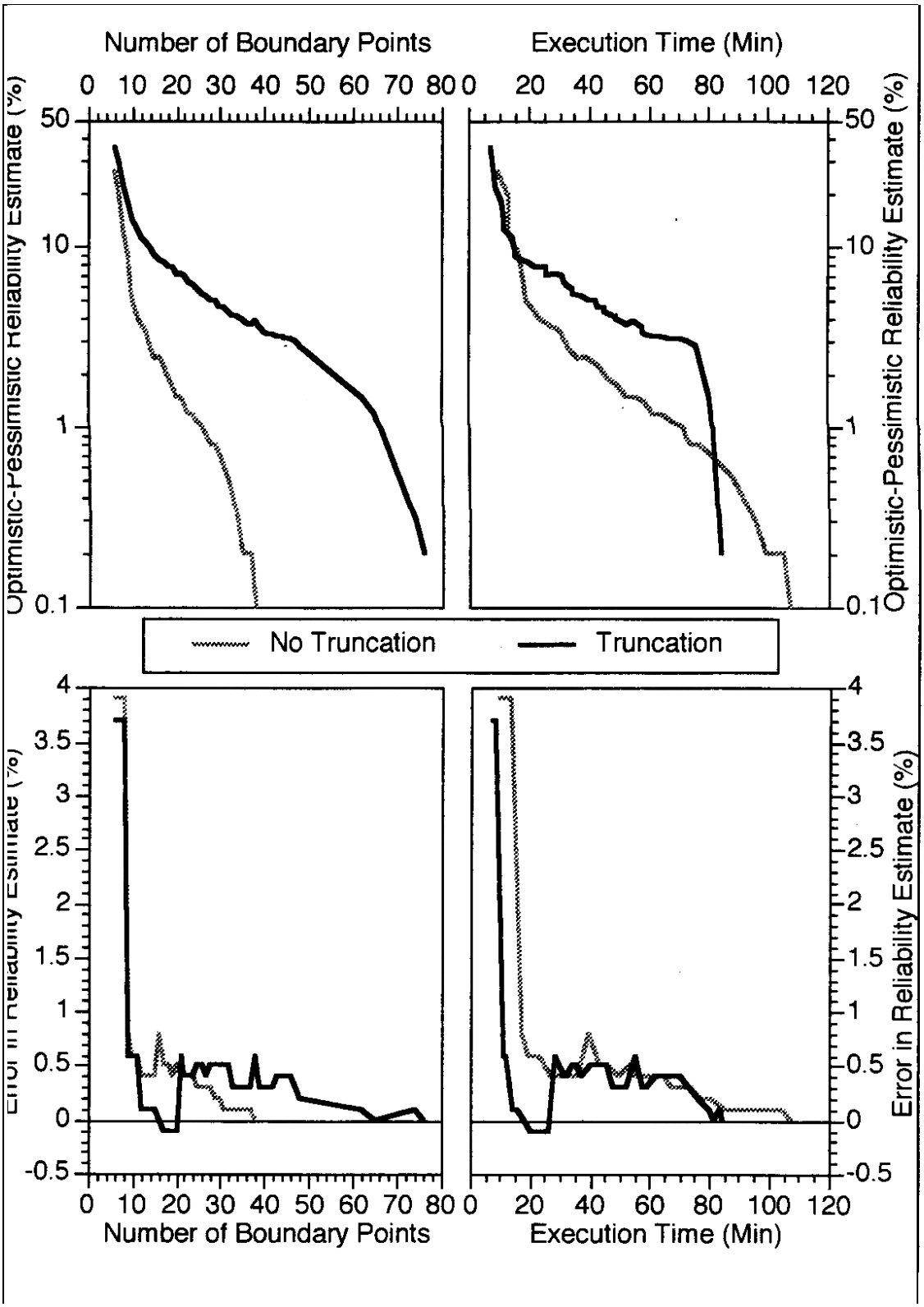

Number of Boundary Points
Execution Time (Min)
Optimistic-Pessimistic Reliability Estimate (%)
0 10 20 30 40 50 60 70 80
0 20 40 60 80 100 120
50
10
1
0.1
No Truncation
Truncation
Error in Reliability Estimate (%)
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
Number of Boundary Points
Execution Time (Min)

Figure V.B.2.b-2
Effect of Truncating the Line-Search in the Tray Efficiency Space

Not-truncating the line-search converges with half as many boundary points but requires slightly more computational effort overall. The biggest difference is towards the end, where truncation speeds up and converges very rapidly, while not-truncating converges at the same rate. This is because the procedure is now resolving individual 'sparse' Monte Carlo points (about 30 remaining, each in a different neighbor group) and truncation provides a boundary point for the cost of just a single simulation, while not truncating requires about p+6 simulations per boundary point. The latter is inefficient.

Different test problems with different parameter dimensions and boundary geometries will undoubtedly provide different comparison results, and truncation is not a critical issue. Also, it must be emphasized that this option is specific to the uniform probability distribution, or any distribution with a fixed minimum or maximum. For this research, therefore, the comparison was not pursued. For all test problems, tray efficiency truncation *was* used.

V.B.3. Thermodynamic Database

This section discusses the general shape of the constraint boundary in the thermodynamic database parameter space. The Modified Wilson solution model was used. Each binary component pair has two uncertain parameters. The boundary tends to be highly elliptical and 'banana-shaped'.

The Modified Wilson equation (Tsuboka and Katayama, 1975) was used for the liquid solution model and thermodynamic database uncertainty representation. The Modified Wilson equation uses two parameters for each binary component pair. For C components, $C(C\text{-}1)$ parameters are required. Each parameter pair is highly correlated. Correlation coefficients are generally -0.99 to -0.999. Each parameter pair is statistically independent of any other parameter pair. So a confidence interval distribution for each parameter pair looks like a very narrow diagonal ellipse, as shown in Figure V.B.3-1. Note that the center is at the designpoint ($\boldsymbol{\theta}^{Des}$), not the origin.

In general, the constraint boundary is diagonally elliptical as well, but it can be highly irregular as shown in the following two figures. Figure V.B.3-2 shows the constraint boundary for the 2MB1-IPM system. Here, the constraint boundary is diagonal, elliptical, and fairly symmetrical. Figure V.B.3-3 shows the constraint boundary for the Acetone-Water system. Here, the constraint boundary is 'banana-shaped' and unbounded in two directions.

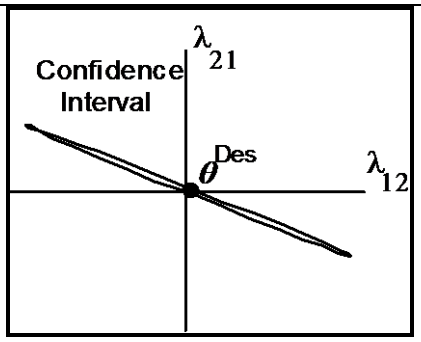


Figure V.B.3-1
Confidence Interval of Database Binary Pair Parameters

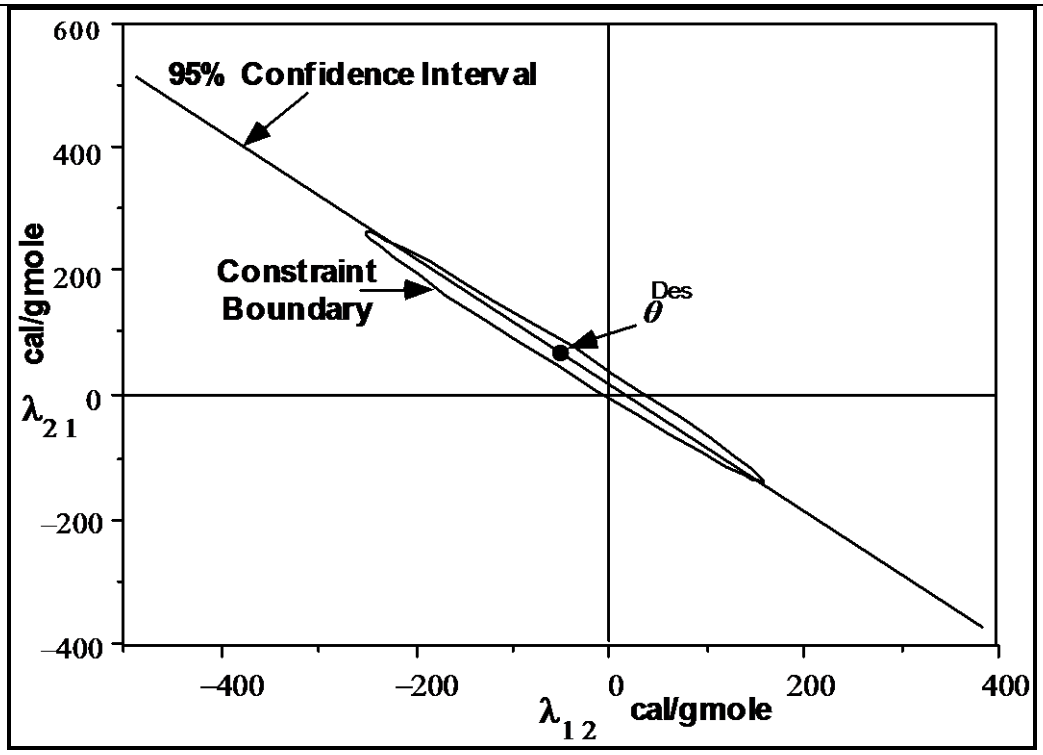


Figure V.B.3-2
λ–Space Constraint Boundary for the 2MB1-IPM Test Case

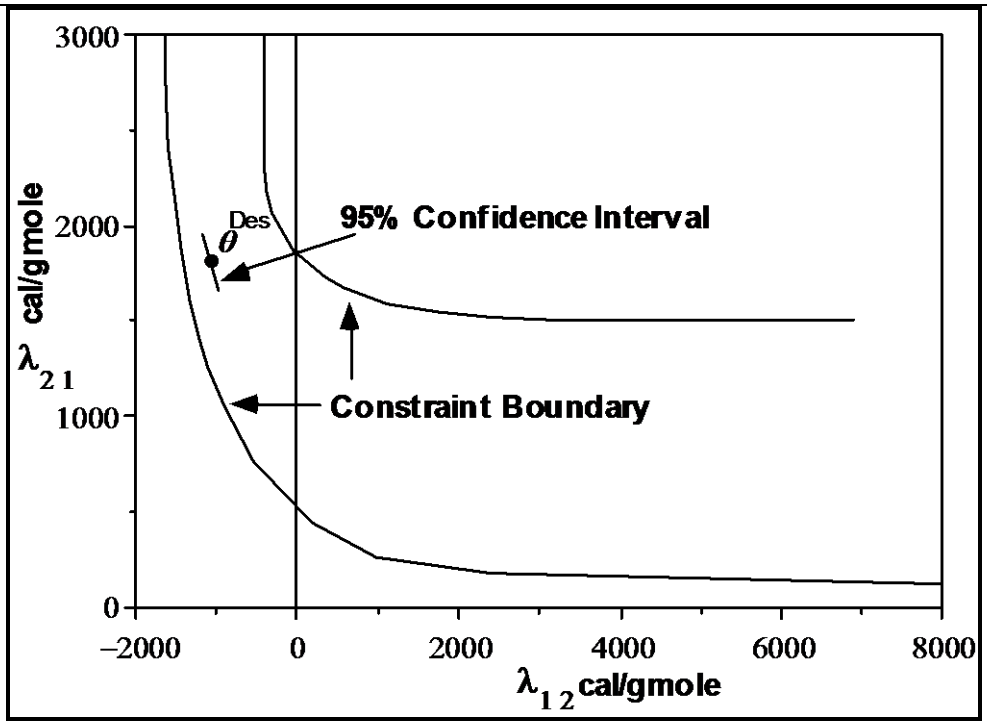


Figure V.B.3-3
λ–Space Constraint Boundary for the Acetone-Water Test Case

V.B.3.a. Including the Database Principal Axis Points to Improve the Initial Boundary-Approximation

Initialization of the boundary-approximation using only the database axial boundary points provides a very poor approximation of the constraint boundary, because the pessimistic boundary is grossly undersized, see Figure V.B.3.a-1. While this error will be corrected, it is more efficient to include the two boundary points along the principal axis of the confidence interval diagonal ellipse with the initial boundary-approximation as shown in Figure V.B.3.a-2. Note that the optimistic and pessimistic boundary-approximations are significantly more accurate.

The improvement in efficiency is due to dimensional considerations. If the database principal axis points are not included, the convergence procedure has to find a boundary point along the database principal axis for *every* hyperquadrant in that direction. A binary component system, for example, requires eight boundary points to achieve the same geometry that two database principal axis points provide, as shown in Figure V.B.3.a-3. For ternary components, forty-eight additional boundary points are required to match just six principal axis points.

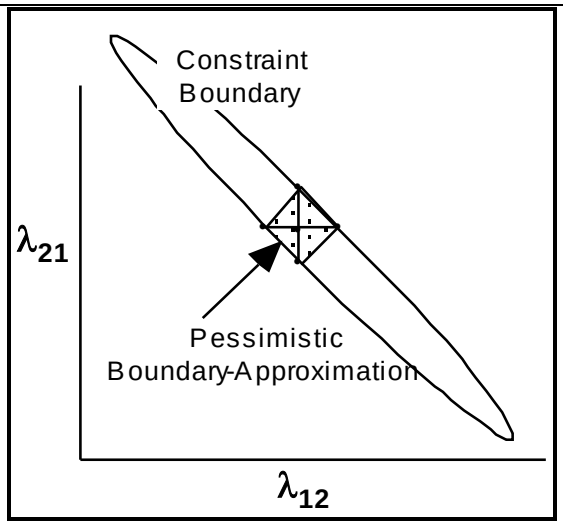


Figure V.B.3.a-1
Axial Boundary Points Alone Provide a Very Poor Pessimistic Boundary-Approximation in the Database Space

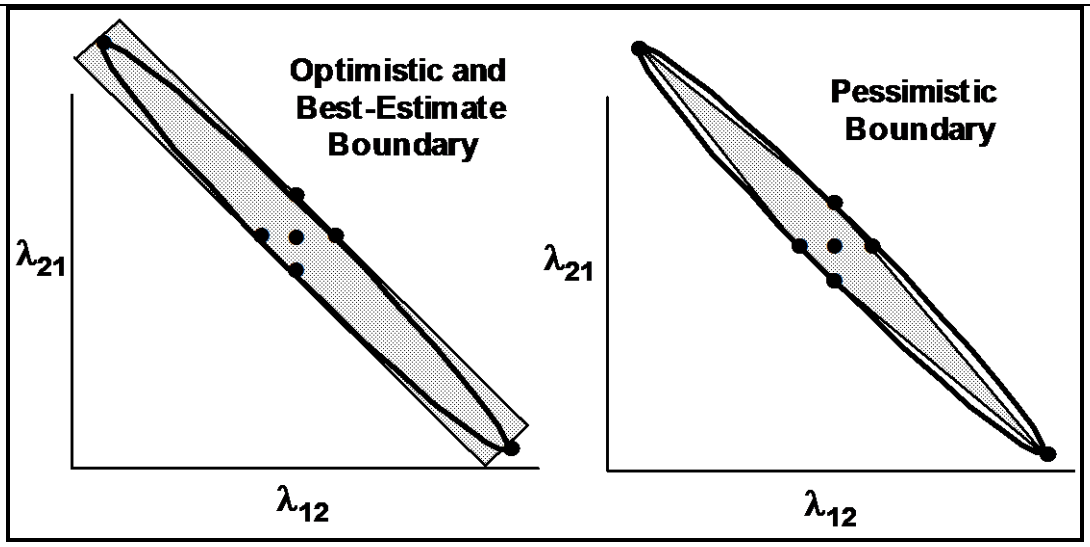


Figure V.B.3.a-2
Initial Boundary-Approximations in the Database Space When Principal Axis Boundary Points are Included

The current boundary-approximation coding requires the database principal axis boundary points to be duplicated with small offsets to provide each hyperquadrant with its own boundary point (see section III.B.1.c and Figure III.B.1.c-2). In other words, for a binary system we first find the two principal axis boundary points then make copies 1 through 8 as shown in Figure V.B.3.a-3. Unfortunately, these offsets may cause slight errors in the boundary-approximation and computer numerical roundoff can become a problem. Better computer coding is ultimately required. For the ternary system, acetone-benzene-toluene, it proved too cumbersome to include their principal axis points at initialization.

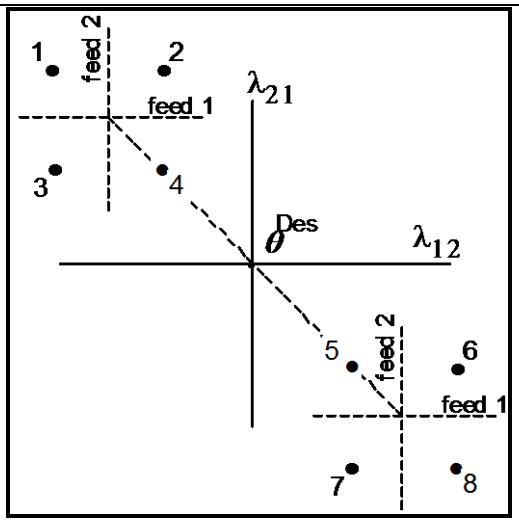


Figure V.B.3.a-3
Dimensional Explosion Causes Boundary Point Duplication Along the Database Principal Axis

V.C. Accuracy and Computational Efficiency of the Boundary-Approximation Procedure for the Distillation Test Problems

This section tests the boundary-approximation procedure with the distillation test problems. Initial and final accuracy is checked in section V.C.1. The computational efficiency is compared with Conventional Monte Carlo in section V.C.2. The boundary-approximation procedure is superior in all aspects.

Figure V.C-1 shows the general convergence characteristics of the boundary-approximation procedure. The important factors are: (1) Initial Accuracy, (2) Convergence Rate, and (3) Final Accuracy. Initial and final accuracy is checked in section V.C.1. Convergence rate, i.e. computational efficiency, is discussed in section V.C.2 which compares the boundary-approximation procedure to Conventional Monte Carlo.

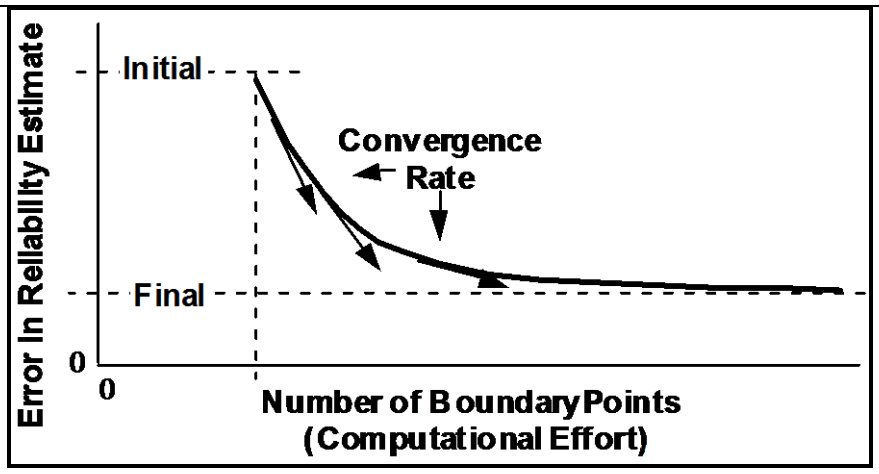


Figure V.C-1. Initial Accuracy, Convergence Rate, and Final Accuracy of the Boundary-Approximation Procedure

V.C.1. Accuracy

What is acceptable accuracy in the reliability estimate? This depends on the reliability and the uncertainty in the statistical descriptions (section II.C.5). If a sensitivity analysis of the latter indicates imprecision of say 5%, there is no need for a reliability estimate within 0.1%. Likewise, a reliability estimate of 60%±10% is sufficiently accurate to reject a proposed design. On the other hand, 96%±3% may not be good enough—the lower bound is seven times more likely to fail than the upper bound.

While no definitive answer is possible, the following sections will show that the question is moot, because the procedure is usually accurate to within 0.5%, which is satisfactory for all but the most rare situations.

V.C.1.a. Initial Accuracy

Accuracy of the reliability estimate at initialization is surprisingly good. Generally, initial accuracy decreases with increasing dimension.

The initial boundary points are the axial points and the database principal axis points. This is sufficient to get the procedure started, however it should not be expected to provide an immediately accurate boundary-approximation and reliability estimate. Figure V.C.1.a-1 compares the initial accuracy for the various test cases as a function of the number of parameters. Except for the acetone-benzene system at higher dimensions, initial accuracy is within 1% which is surprisingly accurate. Generally there is a mild decrease in initial accuracy with increasing dimension. However, the addition of toluene to the acetone-benzene system is a significant exception to this trend.

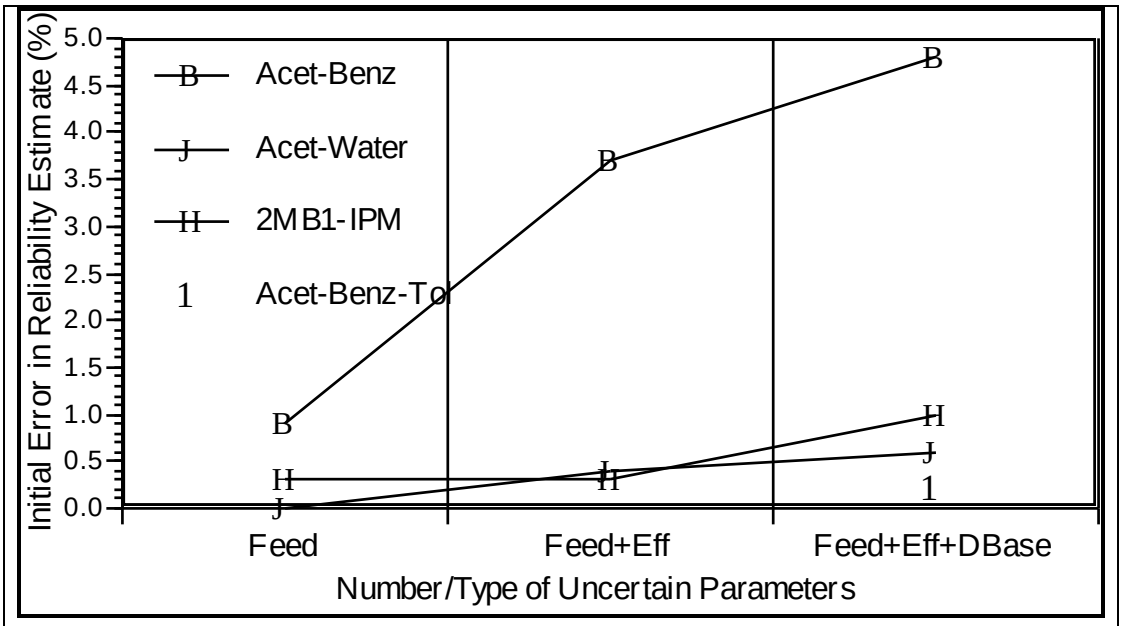


Figure V.C.1.a-1
Initial Accuracy as a Function of the Number of Parameters

Initialization is computationally inexpensive. Figure V.C.1.a-2 plots the equivalent number of Conventional Monte Carlo points that could have been calculated in the same time—between 10 and 60. A 60 point Conventional Monte Carlo run for a 90% reliability design would provide a 95% confidence interval of (-10.5%, +6.2%), which is substantially wider than the typical boundary-approximation error of 1%. Clearly, the boundary-approximation procedure is superior at this point.

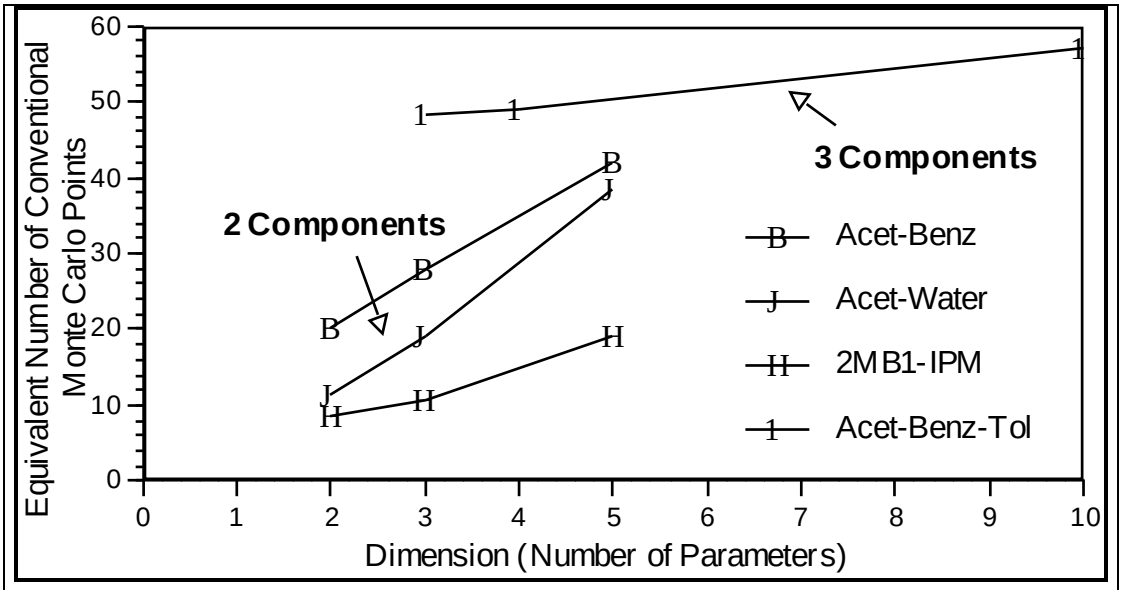


Figure V.C.1.a-2
Equivalent Number of Conventional Monte Carlo Points That Could Have Been Calculated in the Time Required to Initialize the Boundary-Approximation Procedure

V.C.1.b. Final Accuracy

This section looks at the converged or 'final' accuracy of the reliability estimate. Error is generally <0.5%, with a worst case of 1.5% error, which is not significant.

This section looks at the final error of the boundary-approximation. This was obtained by performing a 10^3 point Conventional Monte Carlo simulation, then converging the boundary-approximation with the same random point set. While the Conventional Monte Carlo reliability estimate is not necessarily the 'true' reliability value because of inherent statistical imprecision, it is the correct reliability estimate for that particular random set, so any deviation from this value indicates an error in the boundary-approximation procedure.

The boundary approximation was run for 300 boundary points or until the maximum number of Monte Carlo points in any nearest-neighbor group's region of uncertainty (optimistic-pessimistic) was one. However, this is not sufficient to converge the optimistic and pessimistic boundaries for the highest dimensions, so there are two options for comparison: Tabulate the errors with or without resolution of the Monte Carlo points still within the region of uncertainty. Tabulation without final resolution shows the 'converged' error, or the error that occurs in practice. Tabulation with final resolution shows the 'ultimate' or 'most-final' error, once the procedure has been run to true completion. Table V.C.1.b-1 lists the results for both options.

Table V.C.1.b-1
Final Errors (Bias) in the Boundary-Approximation Reliability Estimate

Test Problem	Parameter Uncertainties	Parameter Dimension	Error (%) *Without* Final Resolution	Error (%) *With* Final Resolution
Acetone				
Benzene	Feed	2	0.0%	0.0%
	Feed+Eff	3	+0.2	0.0
	Feed+Eff+DBase	5	+0.3	0.0
Acetone				
Water	Feed	2	0.0%	0.0%
	Feed+Eff	3	+0.5	0.0
	Feed+Eff+DBase	5	+0.2	0.0
2MB1				
IPM	Feed	2	0.0%	0.0%
	Feed+Eff	3	+0.1	+0.1
	Feed+Eff+DBase	5	+0.6	−0.8
	DBase	2	0.0	0.0
	Eff+DBase	3	0.0	0.0
	Feed+DBase	4	+1.5	+0.1
Acetone **Benzene** **Toluene**	Feed+Eff+DBase	10	−0.3%	−0.4%

Apart from one bad case of 1.5% error (2MB1/IPM – Feed+DBase), final errors are usually less than 0.5%, which is excellent. Note that the errors with resolution are usually lower than errors without resolution, indicating that accuracy will improve if the procedure is fully converged.

The highest errors occurred for the 2MB1-IPM test problem with feed+database uncertainties (with or without tray efficiency). Database uncertainty is highly significant for this case, suggesting that the error may be caused by the complicated geometry of the database parameters. Mathematical transformation of these highly correlated parameters into a more 'regular' space may reduce this error, see section VII.C. This is supported by the observation that as dimension increases, a large percentage of the nearest-neighbor groups have saddle-curvature, which is the most difficult curvature to approximate (section III.B.2.c).

Table V.C.1.b-1 presents the *net* error or bias. In fact, the boundary-approximation procedure may have positive and/or negative errors. Some points which were inside the boundary-approximation might actually be outside the true boundary, and vice-versa.

Generally, these positive and negative errors will cancel to some extent—the net error. Thus, while the net error may be low, the boundary-approximation could actually be highly inaccurate. To check this, Table V.C.1.b-2 presents the same results as Table V.C.1.b-1, except *total* errors are reported, i.e., the sum of the positive and negative errors (absolute values). While these errors are naturally higher, they remain acceptably small.

Table V.C.1.b-2
Final Errors (Total) in the Boundary-Approximation Reliability Estimate

Test Problem	Parameter Uncertainties	Parameter Dimension	Error (%) *Without* Final Resolution	Error (%) *With* Final Resolution
Acetone Benzene	Feed	2	0.0%	0.0%
	Feed+Eff	3	0.4	0.0
	Feed+Eff+DBase	5	0.5	0.0
Acetone Water	Feed	2	0.0%	0.0%
	Feed+Eff	3	0.5	0.0
	Feed+Eff+DBase	5	0.2	0.0
2MB1 IPM	Feed	2	0.2%	0.2%
	Feed+Eff	3	0.1	0.1
	Feed+Eff+DBase	5	2.2	1.4
	DBase	2	0.0	0.0
	Eff+DBase	3	0.0	0.0
	Feed+DBase	4	1.7	0.3
Acetone Benzene Toluene	Feed+Eff+DBase	10	0.5%	0.4%

V.C.2. Computational Efficiency—Comparison with Conventional Monte Carlo

The computational efficiency of the boundary-approximation procedure is compared to Conventional Monte Carlo: The boundary-approximation procedure is more efficient in all cases.

The previous sections showed that initial and final accuracy of the boundary-approximation procedure is very good. This section looks at the convergence rate in comparison to Conventional Monte Carlo. This comparison is difficult, however, because of

their inherently different statistical properties: Conventional Monte Carlo is accurate (unbiased), but imprecise. The boundary-approximation is arbitrarily precise, but of uncertain accuracy. Schematically, this is shown in Figure V.C.2-1.

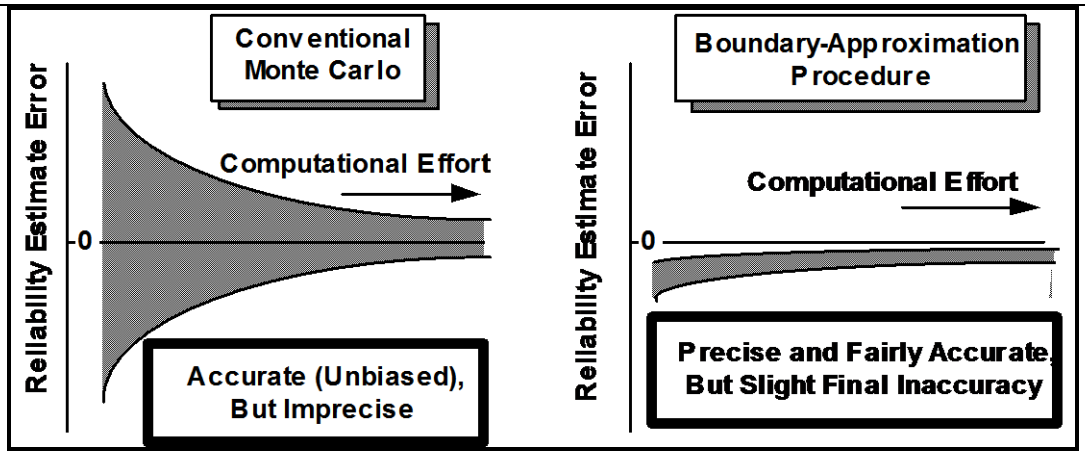


Figure V.C.2-1
Accuracy and Precision Comparison of the Conventional Monte Carlo and Boundary-Approximation Procedures

Error for the boundary-approximation procedure is the difference between the reliability estimate and the true value. Error for Conventional Monte Carlo is a statistical quantity, such as a 95% confidence interval. Therefore, a comparison between the two is necessarily statistical.

As discussed in the previous section, Conventional Monte Carlo was first used to calculate the 'true' reliability for a 10^3 random number set. Then, the boundary-approximation was converged with this same random set. The difference between the converging reliability estimate and the reliability estimate from Conventional Monte Carlo is an estimate of the error of the boundary-approximation procedure at that point.

Next, the time taken by the boundary-approximation procedure was converted to the equivalent number of Conventional Monte Carlo points that could have been calculated. This allows Conventional Monte Carlo confidence intervals to be calculated using statistical tables (Diem and Lentner, 1975) or estimation equations (Appendix B, Eq B-2-12). The width of these confidence intervals may then be compared with the error of the boundary-approximation procedure.

The following three sections present the results. Section V.C.2.a compares reliability estimate accuracy as a function of the computational effort. Section V.C.2.b reverses the axes and plots the time-ratio (time required for Conventional Monte Carlo divided by the time required for the boundary-approximation) as a function of accuracy in the reliability estimate.

Section V.C.2.c plots the statistical odds that the boundary-approximation procedure is more accurate than a Conventional Monte Carlo run. In all cases, the boundary-approximation is easily superior.

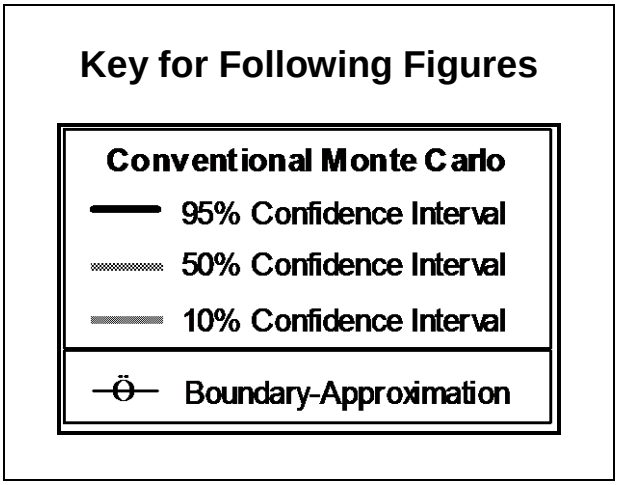
V.C.2.a. Convergence Profiles: Accuracy as a Function of Computational Effort

This section compares the error of the boundary-approximation reliability estimate with the Conventional Monte Carlo confidence intervals as a function of computational effort. Computational effort is shown as the equivalent number of Monte Carlo points that could have been calculated in that time. Figure V.C.2.a-1 plots the comparison for acetone-benzene. Figure V.C.2.a-2 plots the comparison for acetone-water. Figure V.C.2.a-3 plots the comparison for 2MB1-IPM. Figure V.C.2.a-4 plots the comparison for acetone-benzene with and without toluene. 95%, 50%, and 10% confidence intervals are shown. The 95% confidence interval is the standard error estimation quantity. The 50% confidence interval is the median accuracy of Conventional Monte Carlo—half the runs are more accurate, half are less accurate. The 10% confidence interval is a reasonable 'best-case' for Conventional Monte Carlo—out of 100 runs, this is the 10'th best. Note that in many of these plots, only the 10% confidence interval is visible—the 50% and 95% confidence intervals are so wide they do not show up on the scale. For conversion, the 50% (95%) confidence interval is 5.37 (15.6) times wider than the 10% confidence interval.

For acetone-benzene, the high initial error (1%-5%) in the boundary-approximation procedure is rapidly reduced to within 0.5%. Small oscillations then predominate until final convergence which takes substantially longer. Except for the initial error, accuracy is approximately equal to the 10% confidence interval of the Conventional Monte Carlo.

For acetone-water, initial error is low (<1%) so there is no significant initial error correction. Once again, accuracy is approximately equal to the 10% confidence interval of the Conventional Monte Carlo.

For 2MB1-IPM, convergence for the feed-only and feed+tray.efficiency cases is rapid and smooth. However, the full dimensional case (for which database uncertainty is statistically very significant) is noisy and has difficulty converging. Accuracy is approximately equal to the 10% confidence interval of the Conventional Monte Carlo, except at convergence for the full dimensional



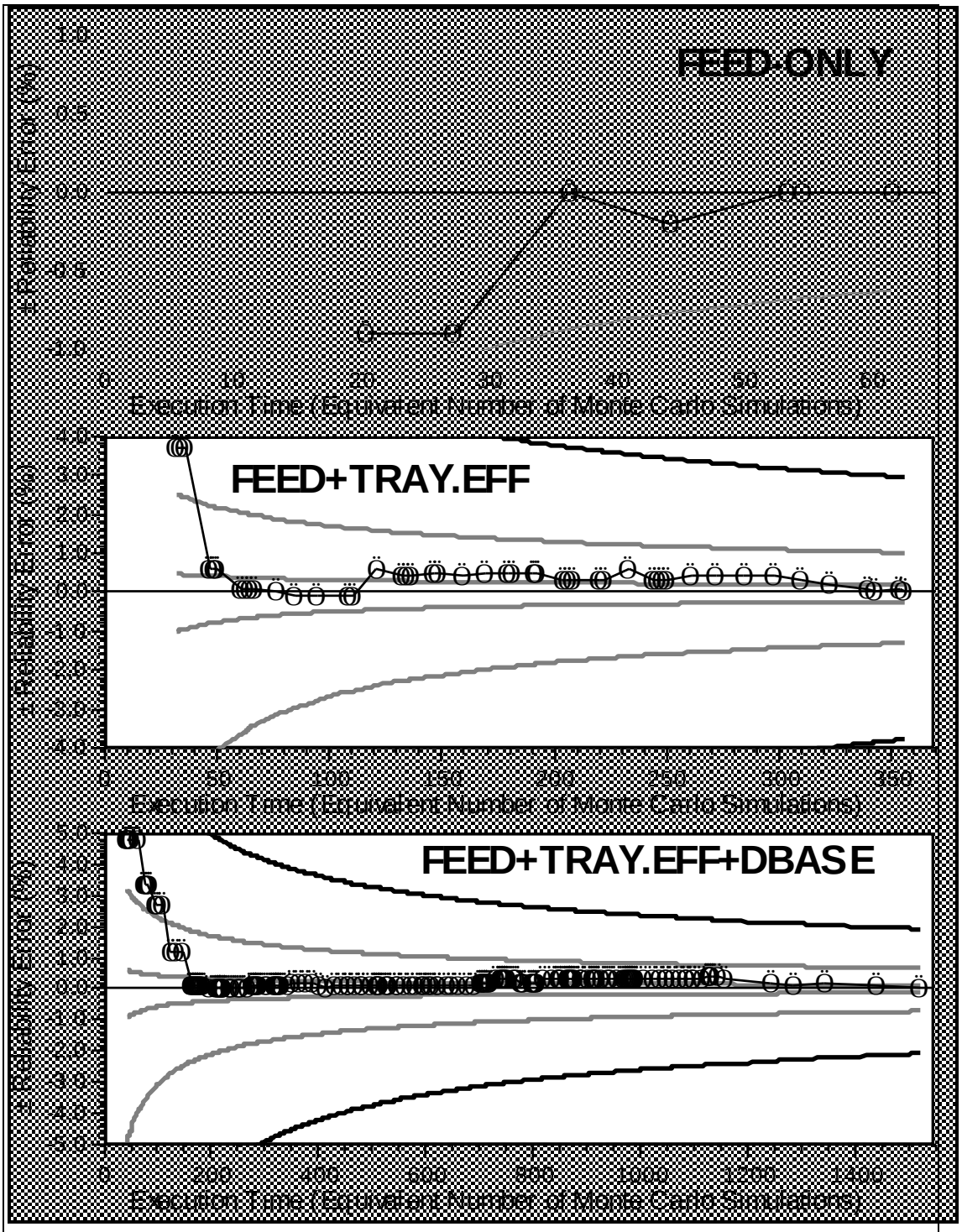


Figure V.C.2.a-1
Comparison of Boundary-Approximation and Conventional Monte Carlo: Accuracy as a Function of Computational Effort—Acetone-Benzene

FEED-ONLY is immediately accurate

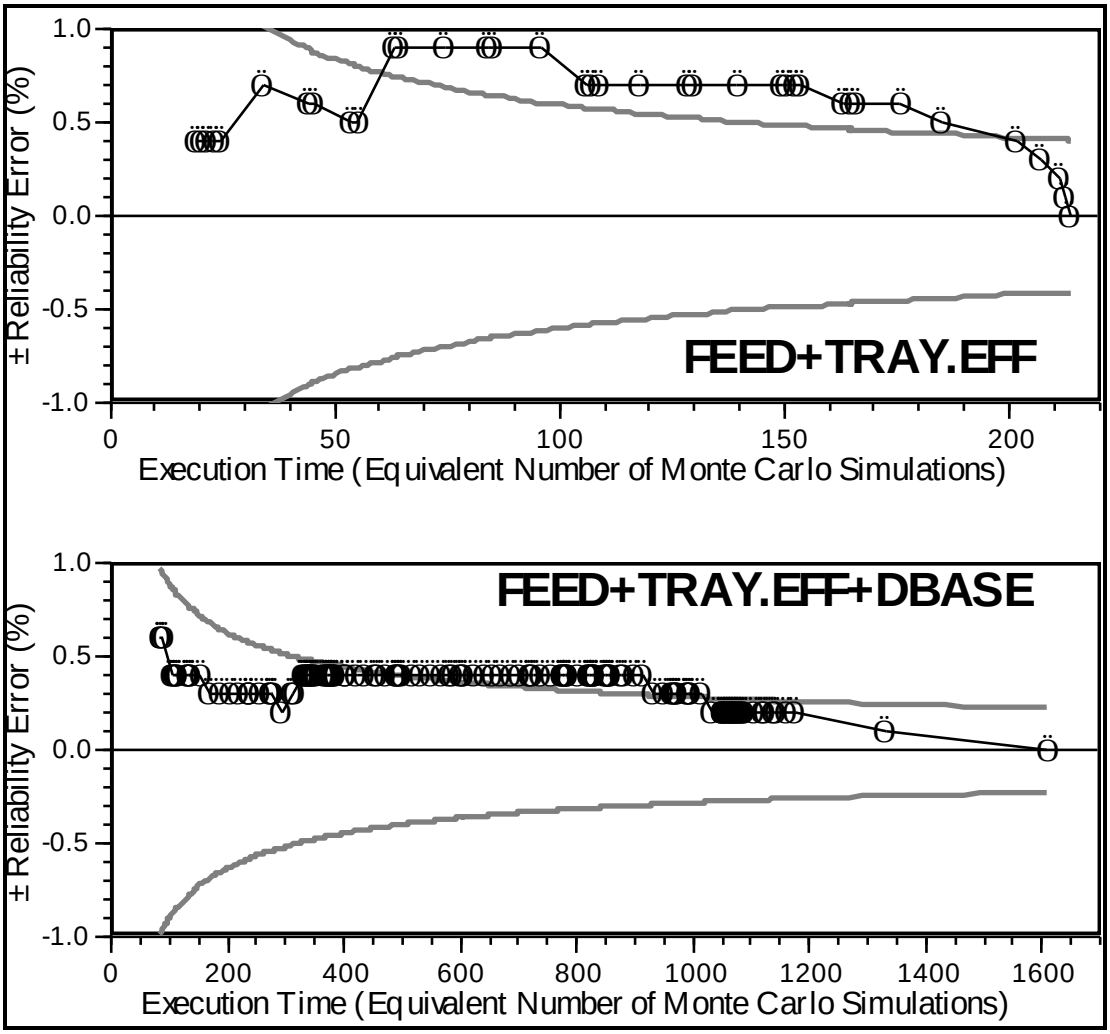


Figure V.C.2.a-2
Comparison of Boundary-Approximation and Conventional Monte Carlo: Accuracy as a Function of Computational Effort—Acetone-Water

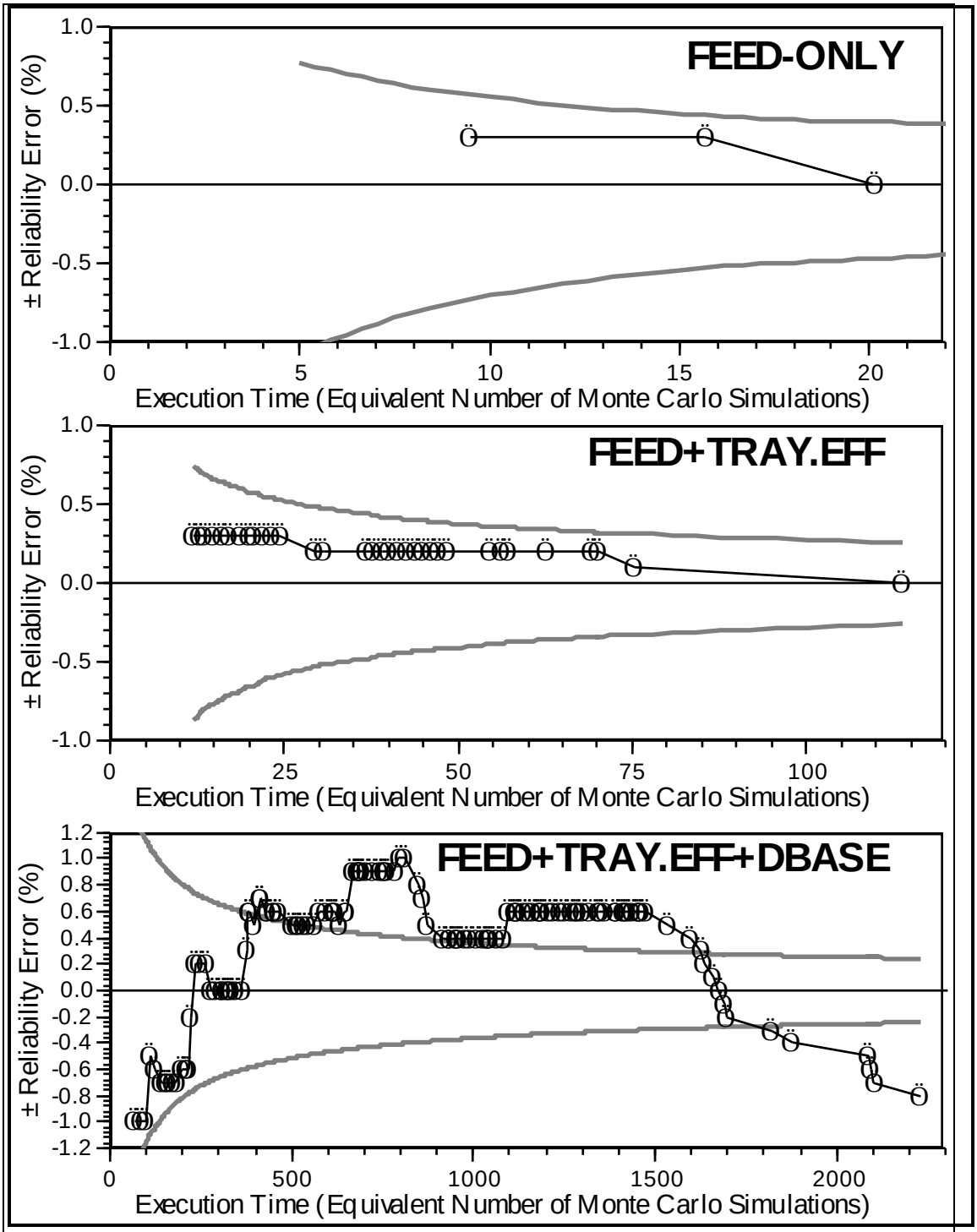


Figure V.C.2.a-3
Comparison of Boundary-Approximation and Conventional Monte Carlo: Accuracy as a Function of Computational Effort—2MB1-IPM

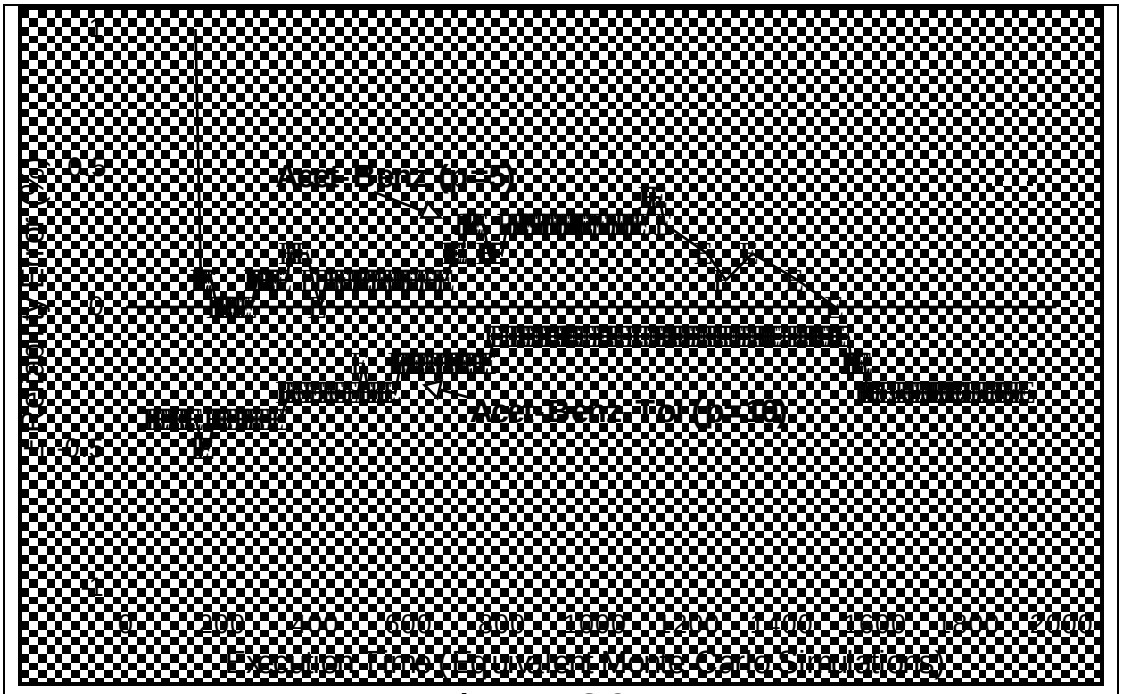


Figure V.C.2.a-4
Comparison of Boundary-Approximation and Conventional Monte Carlo: Accuracy as a Function of Computational Effort—Acetone-Benzene + Toluene

case where final error is 0.8% which is significantly higher than the 10% confidence interval.

For acetone-benzene-toluene, accuracy is approximately equal to the 10% confidence interval of the Conventional Monte Carlo, except at the final stages where error increases to 0.3% (the procedure had not converged after 300 boundary points—final error is 0.4%). Note that the increased dimension is not a significant factor.

The boundary-approximation procedure quickly corrects initially poor reliability estimates (>1.0% error) as shown by the acetone-benzene test problem with higher dimensions. This correction does not occur when initial error is low. For all test problems, any error is quickly reduced to within 0.5%-1.0%. Thereafter, the procedure may oscillate within ±0.5%-1.0% until final convergence, which takes a significant time for the higher dimensions.

Clearly, the boundary-approximation procedure converges with about the same computational efficiency as the Conventional Monte Carlo 10% confidence interval, which is excellent. The only problem with the boundary-approximation procedure is its potential for slight final error at convergence. Even then, however, the equivalent error from Conventional Monte Carlo is likely to be significantly greater.

V.C.2.a.i. Accuracy After 100 and 500 Equivalent Monte Carlo Simulations

Since 100 and 500 Conventional Monte Carlo simulations are reasonably acceptable computational efforts, this section compares the accuracy that both procedures will provide after this amount of time. After 100 simulations, the 95% confidence interval for Conventional Monte Carlo is about ±5% to ±17%. In contrast, the boundary-approximation procedure had already converged for the lower dimension test problems and was fairly accurate for the higher dimension test cases (usually <1% with a worst-case error of 2.7%). After 500 simulations, the 95% confidence interval for Conventional Monte Carlo is about ±3% to ±8%, while the worst-case error for the boundary-approximation procedure was 0.5%. Clearly, the boundary-approximation procedure is much more accurate.

Table V.C.2.a.i-1 lists the ratio of the accuracies of the two procedures. Except for the anomaly of the acetone-benzene full dimensional case which had high error after 100 simulations (2.7%), the boundary-approximation procedure is about 10-100 times more accurate than the 95% confidence interval of the equivalent Conventional Monte Carlo procedure.

Table V.C.2.a.i
Width of the 95% Confidence Interval of the Conventional Monte Carlo (Max%-Min%)/2 Divided by the Error of the Boundary-Approximation Procedure After 100 and 500 Equivalent Monte Carlo Simulations

100 Simulations	Acet-Benz	Acet-Water	2MB1-IPM	Acet-Benz-Tol
Feed	∞	∞	∞	
Feed+Eff	64	11	128	
Feed+Eff+DBase	2.9	29	18	15
500 Simulations				
Feed	∞	∞	∞	
Feed+Eff	∞	∞	∞	
Feed+Eff+DBase	35	15	16	10

V.C.2.b. Execution Time-Ratio (Conventional Monte Carlo / Boundary-Approximation) as a Function of Accuracy in the Reliability Estimate

This section reverses the axes of the plots in the previous section and shows execution time as a function of accuracy in the reliability. Computational effort is expressed as a *time-ratio*, i.e., the time required by Conventional Monte Carlo divided by the time required by the boundary-approximation procedure to obtain equivalent accuracies. A value greater than one

indicates that the boundary-approximation procedure is superior. The time-ratio for the 95% Conventional Monte Carlo confidence intervals is plotted in Figure V.C.2.b-1 for all test problems. (For clarity, Appendix E-1 provides the same plot repeated for each test problem.) To convert the time-ratio to 50% and 10% confidence intervals, divide the y-axis value by 8-9 and 230-290 respectively (the value is inexact). Note: Since the confidence intervals are not symmetrical, (Max%-Min%)/2 was used instead. Also, because of the noisy convergence and number of points, for clarity only minimum and maximum time-ratios are plotted for each level of accuracy.

For reliability estimate errors of 1%-5%, the time-ratio is only 3-10. However, the only time this occurred (acetone-benzene) the initial error was quickly corrected to less than 1%. Where error is less than 1%, the boundary-approximation procedure is 20-7000 times faster than the 95% Conventional Monte Carlo confidence interval. The time-ratio increases as accuracy increases, primarily because the boundary-approximation procedure converges much more rapidly than Conventional Monte Carlo.

In Figure V.C.2.b-1, the highest dimensional cases (feed+tray.efficiency+database) are shown with solid symbols, while the low dimensional cases are shown with hollow symbols. The range of values is greater than any trend. While the highest dimensional case (acetone-benzene-toluene) has the lowest time-ratio and time-ratio decreases with dimension for the acetone-benzene case, dimension has no effect for acetone-water and 2MB1-IPM. Thus we may conclude that dimension is not a major factor for the dimensional range studied.

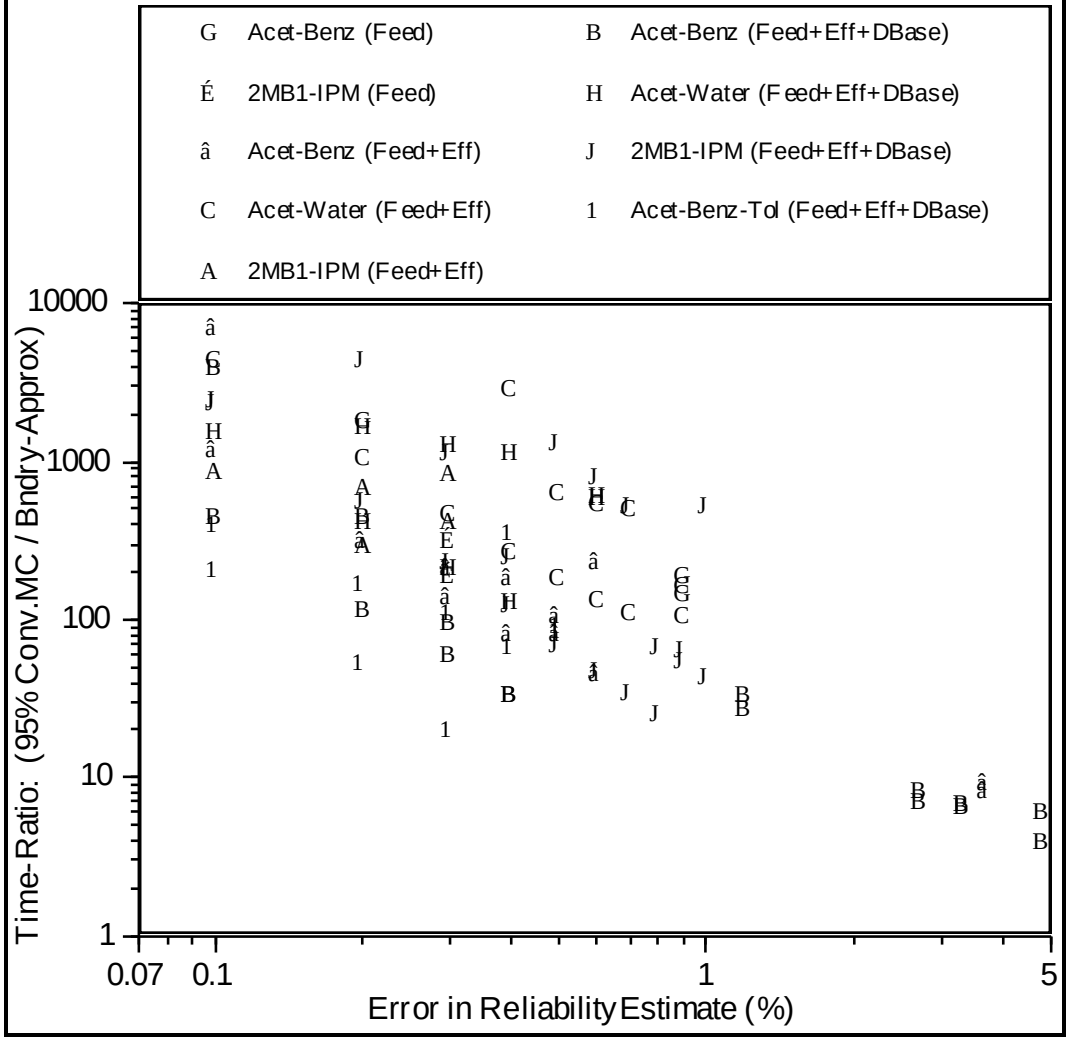


Figure V.C.2.b-1
Time Required for 95% Confidence Interval of Conventional Monte Carlo Divided by Time Required for Boundary-Approximation Procedure as a Function of Reliability Estimate Accuracy

V.C.2.c. Statistical Odds that the Boundary-Approximation Procedure Is More Accurate than Conventional Monte Carlo

This section calculates the statistical odds that the boundary-approximation procedure is more accurate than Conventional Monte Carlo for the same computational effort. Details of this calculation are provided in Appendix B-3. Figure V.C.2.c-1 shows the odds as a function of computational effort (in units of equivalent Conventional Monte Carlo points) for all test problems combined. Since the individual plots are very noisy, only the general trend is shown by the shaded region. The arrows show the general areas for each type of test problem,

although there was much overlap. Appendix E-2 provides the detailed plot for each test problem.

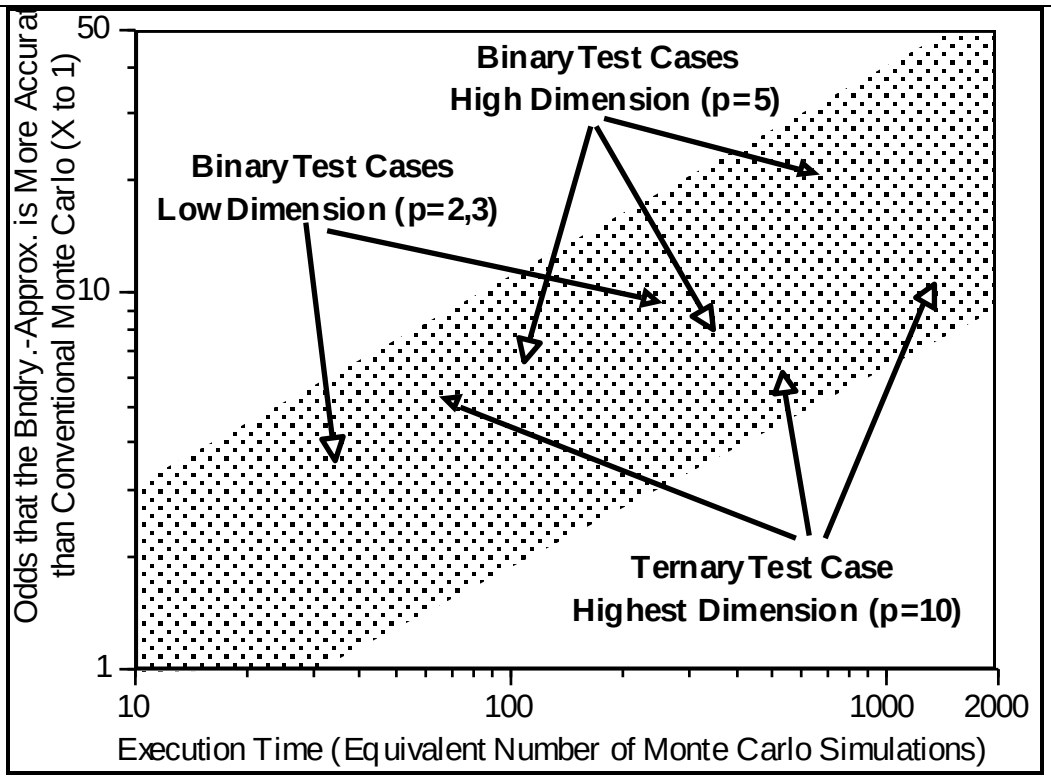


Figure V.C.2.c-1
Likelihood that the Boundary-Approximation Procedure is More Accurate than the Conventional Monte Carlo Procedure as a Function of Computational Effort

The larger the dimension, the more effort is required to initialize and converge the boundary-approximation procedure. Nevertheless, the superiority of the boundary-approximation procedure is fairly independent of dimension, although it does decrease for the highest dimension case of acetone-benzene-toluene (still easily superior, however).

For the equivalent of 100 Conventional Monte Carlo simulations, the boundary-approximation procedure is about 2 to 10 times more likely to be accurate. At 500 equivalent simulations, it is about 4 to 30 times more likely to be accurate The improvement with computational effort is an indication that the boundary-approximation converges faster than Conventional Monte Carlo and any final error of the boundary-approximation is comparatively insignificant.

V.C.3. Effect of Parameter Dimension on the Boundary-Approximation Procedure Convergence Rate

This section examines the effect of parameter dimension of the boundary-approximation procedure convergence rate. Qualitatively, each additional parameter should halve the convergence rate. However, this is not always observed and further testing with geometrical test problems is required. Also, initial accuracy is so good that convergence is not critical. Parameter dimension is not an insurmountable problem.

In the previous sections, it was seen that the boundary-approximation procedure is orders of magnitude superior to Conventional Monte Carlo independent of parameter dimension for the test problems studied. However, the maximum dimension was only three chemical components and $p=10$. Our knowledge of the calculations tells us that the boundary-approximation procedure becomes more complex as dimension increases (Conventional Monte Carlo is independent of dimension). Of particular importance is that the number of nearest-neighbor groups (section III.B.1) doubles with each additional parameter, and therefore twice as many additional boundary points are required to 'fill in the gaps'. Qualitatively therefore, each additional parameter should halve the convergence rate and at some point the superiority of the boundary-approximation procedure may disappear.

To properly study the effect of parameter dimension, geometrical test problems should be used instead of distillation. This is because statistically significant and insignificant parameters and/or linear or nonlinear geometries will have different effects. Distillation is too complicated a model to control these factors properly, and the computational requirements are unnecessarily high.

However, it is possible to perform some preliminary studies to predict what will occur as dimension increases. This section attempts to do so. Unfortunately, we cannot simply extrapolate the increasing dimensional sequence—feed-only ($p=2$), feed+tray.efficiency ($p=3$), feed+tray.efficiency+database ($p=5$)—because the geometrical variations are too radical. Adding toluene to acetone-benzene was an attempt to minimize this. However, this can only show the effect of adding statistically *insignificant* parameters. And other factors are unavoidably altered, which fogs the comparison. See Appendix D for an idea of the wide variety of convergence rates encountered.

Instead, this section converges the full dimensional test problems (feed+tray.efficiency+database) with and without the assumption of symmetry in the database parameter uncertainties. By assuming symmetry, the dimension of the binary test problems can be halved and the ternary test problem can be reduced by a factor of eight, with a minimum of disruption to other factors. If the hypothesis is correct, the binary test problems should converge twice as fast and the ternary test problem should converge eight times faster when symmetry is assumed.

2MB1-IPM was run to look at highly significant parameters. Acetone-benzene and acetone-benzene-toluene were run to look at less significant parameters at different dimensions. Figure V.C.3-1 plots the convergence comparison for acetone-benzene. Figure V.C.3-2 plots the convergence comparison for 2MB1-IPM. Figure V.C.3-3 plots the convergence comparison for acetone-benzene-toluene. The y-axis is the absolute value of the difference between the current reliability estimate and the converged value. The x-axis is the number of boundary points divided by a factor (x), where x is 1 or 2 as noted. 10^5 point Monte Carlo integration was used to improve resolution.

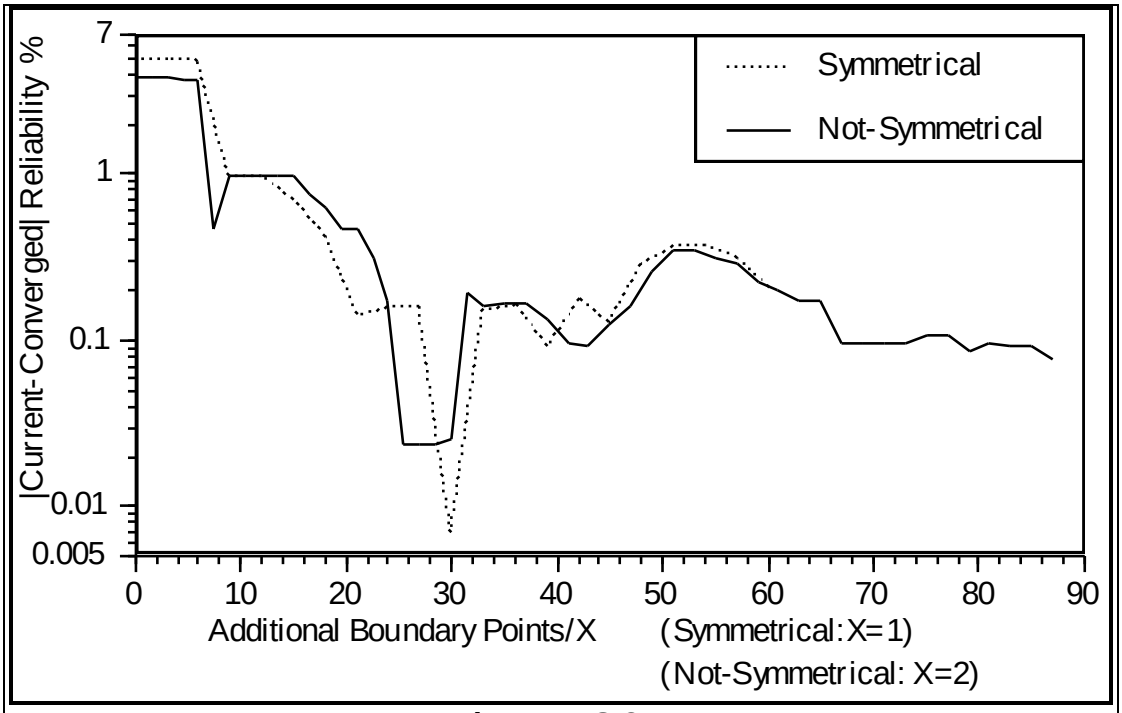


Figure V.C.3-1
Convergence Comparison of Acetone-Benzene With and Without Symmetry in the Database Parameters

For acetone-benzene, the unsymmetrical convergence requires twice as many boundary points to converge with the same profile, which supports the hypothesis. 2MB1-IPM also supports the hypothesis, although the unsymmetrical convergence is noisier due to the highly nonlinear geometry.

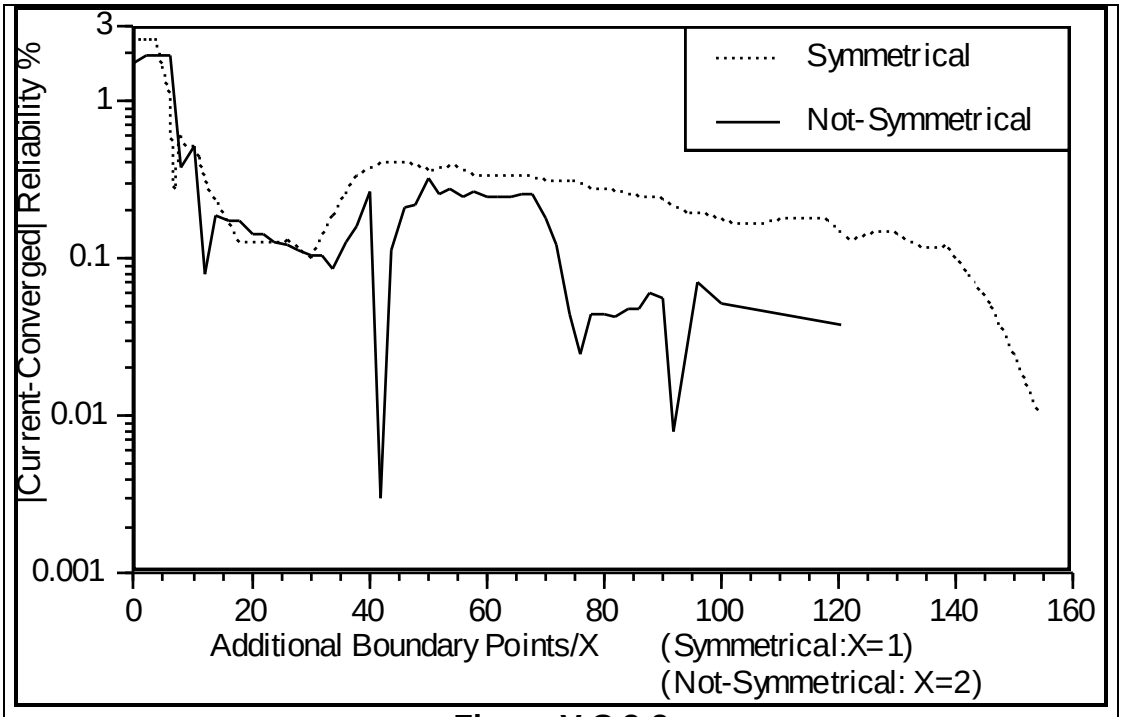


Figure V.C.3-2
Convergence Comparison of 2MB1-IPM With and Without Symmetry in the Database Parameters

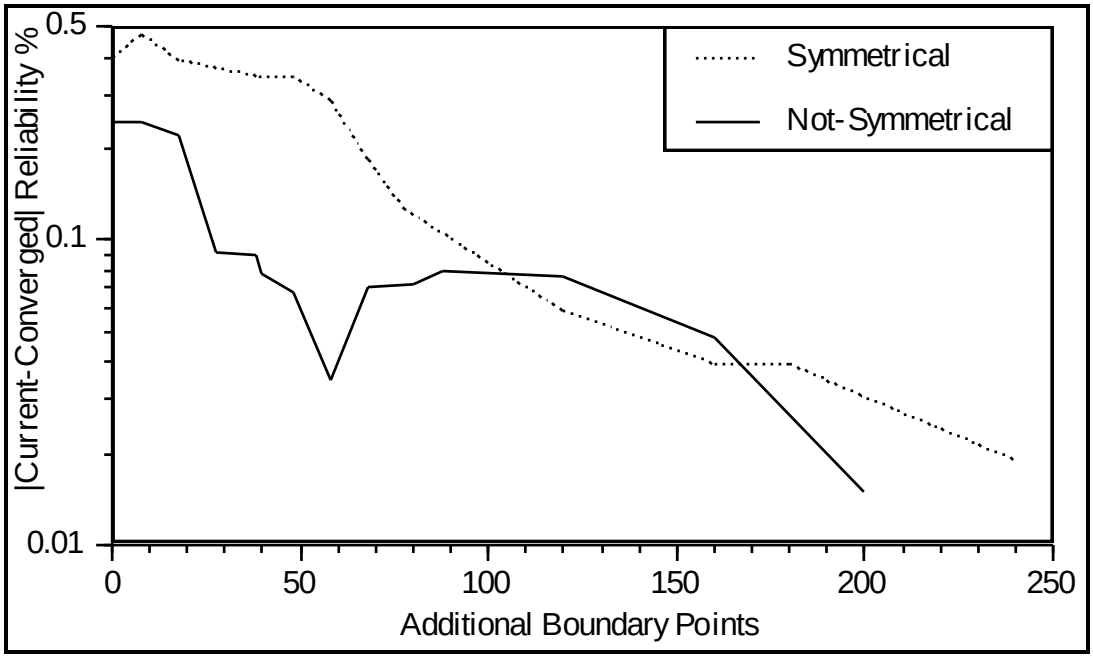


Figure V.C.3-3
Convergence Comparison of Acetone-Benzene-Toluene With and Without Symmetry in the Database Parameters

For the acetone-benzene-toluene case, Figure V.C.3-3 shows roughly *equivalent* convergence rates (same slopes) with and without symmetry instead of the expected eight-fold difference. Thus, the hypothesis is contradicted. However, this picture may be misleading because the initial reliability estimate is very accurate (<0.5%), so the convergence procedure exhibits no initial systematic correction. In contrast, the two other cases have initial errors of 2%-5% and consequently show very strong initial corrections.[1] Therefore, we may simply be observing initial oscillations in the reliability estimate. Also, at this highest dimension neither run has converged, so perhaps we are not seeing enough of the plot to make a valid comparison. Nevertheless, we cannot conclude from this test that each additional parameter always halves the convergence rate. Testing with geometrical problems is required for further study.

While the results of this section are disappointing if our goal was to estimate the dimension where the boundary-approximation procedure loses its superiority, we may observe to the contrary: (1) If the initial reliability estimate of the boundary-approximation procedure is accurate, dimension is irrelevant. (2) And on those hopefully rare occasions when the initial reliability estimate is poor *and* dimension is very high, the initial error will be corrected immediately—albeit slowly. Also, it may be possible to accelerate this phase as discussed in section VII.E**.**

[1] It was observed in section V.C.2.a. that the procedure quickly corrects any gross initial error (>1.0%), but may oscillate within 0.5%-1.0% once this is accomplished.

VI. Conclusions & Recommendations

Summary

- This research has removed the computational barriers for estimating design reliability. Accurate and precise reliability estimates can now be obtained for orders of magnitude less computational effort than before. Calculations that would have taken days and weeks on a microcomputer can now be completed within hours.
- Work should continue along two parallel paths:
  (1) The boundary-approximation procedure should be extended to hybrid (probability+possibility) statistics. The procedure should be optimized and tested with higher dimensions, removing some minor limitations. Explicit handling of control variables should be developed. Optimum handling of multi-equipment flowsheets should be studied.
  (2) More accurate statistical descriptions of the process design uncertainties must be developed.

Comparison Between Boundary-Approximation Procedure and Conventional Monte Carlo Procedure

Since this research was limited to probability-statistics testing, the basis for comparison was Conventional Monte Carlo:

- For the same level of accuracy, the boundary-approximation procedure is 20-7000 times faster than the 95% confidence interval of Conventional Monte Carlo.
- For the same computational effort, the boundary-approximation procedure is 10-100 times more accurate than the 95% confidence interval of Conventional Monte Carlo.
- For the equivalent computational effort of 100 Conventional Monte Carlo simulations, the boundary-approximation procedure is 2-10 times more likely to be accurate than Conventional Monte Carlo. After 500 equivalent simulations, it is 4-30 times more likely to be accurate.
- The boundary-approximation procedure requires a small initialization effort equivalent to 10-60 Conventional Monte Carlo simulations (increasing with the number of uncertain parameters). Initial accuracy is usually ±1% and always ±5%, which is superior to the equivalent Conventional Monte Carlo. Convergence is much faster than Conventional Monte Carlo. Final accuracy is usually 0.0% – 0.5%, with a worst case of 1.5%.

Disadvantages of the Boundary-Approximation Procedure

- The reliability estimate may be slightly inaccurate—usually <0.5%; worst case=1.5%.
- Computational effort increases and convergence rate should decrease as parameter dimension increases.

Recommendations

Improvements to Boundary-Approximation Procedure

General

- The procedure should be extended to hybrid statistics. This requires no modification to the mathematical boundary-approximation. However, the statistical calculation of design reliability and the selection of the next boundary point require modification.
- The procedure should be extended to handle control variables explicitly without substitution.
- More advanced mathematical boundary-approximations should be developed. This will improve accuracy and computational efficiency.
- For large parameter dimensions, some parts of the computer code become slow and require a great deal of memory. This can be minimized by coding modifications that should be developed.
- The current algorithm cannot handle boundary points where any parameters are at their design values. This needs to be corrected.
- Special coding should be developed to optimize handling of 'bounded' uncertain parameters, i.e., that have a fixed maximum and/or minimum.

Improving Computational Efficiency

- Since the nested-loop algorithm is fairly efficient (approximately $6+p$ simulations per boundary point), and has the significant advantage that no simulation modification is required, it is sufficient for the near future. The simultaneous algorithm may be considered for large dimension problems and for explicit handling of control variables.
- The nested-loop algorithm can be improved by more advanced line-searching algorithms.
- Convergence tolerances should be optimized for significant computational improvement.

Parameter Screening

- To minimize dimension, a screening procedure should be developed to identify the significant parameters, see section VII.A.

Large Dimension Testing with Geometrical Models

- The number and type of parameters are the biggest factors affecting the boundary-approximation procedure. Their exact influence is better studied with geometrical boundaries. Questions to resolve include:
  (1) How many dimensions can the boundary-approximation procedure comfortably handle?
  (2) Is there any difference in the effect of statistically significant and insignificant parameters?
  (3) Is there any difference in the effect of linear and nonlinear parameters?

Distillation Process Model Development

- The distillation model contains additional uncertainties that should be included, such as flood point, tray/packing efficiency models, pressure drop, vapor phase thermodynamic models, and heat exchangers.
- Only normal and uniform probability distributions were used in testing. Research is required to determine which statistical distributions are most accurate.
- Parameter transformations may reduce the nonlinearity of the constraint boundary and therefore improve the boundary-approximation. For example, the parameters used in the Modified Wilson liquid solution model (Tsuboka and Katayama, 1975) are highly pairwise correlated. These parameters should probably be transformed to more 'regular' coordinates. This is discussed in section VII.C.
- Different combinations of control variable substitutions (see section IV.B.3) should be tested to see which provides the most favorable geometry.
- The boundary-approximation procedure should be used to study the effectiveness of various types of safety factors for distillation design. Perhaps simple heuristics can be developed and/or confirmed for preliminary design applications.

General Model Development

- Additional process operation models should be tested to determine what constraint boundary shapes occur in practice.
- Extension to entire process flowsheets (multiple unit operations) should be studied.

VII. Some Suggestions For Future Work

This section discusses some potential improvements to the boundary-approximation procedure that are not based on actual research, but are important nevertheless. Discussion is necessarily qualitative and conceptual since the details are unknown.

The boundary-approximation procedure has two principal limitations: (1) Increasing computational costs as parameter dimension increases, and (2) Potential for slight inaccuracy in the reliability estimate. This section suggests some ways to minimize and eliminate these problems. Parameter dimension can be handled by parameter screening (section VII.A) and better geometric treatment (sections VII.B, VII.C, and VII.D). Reliability errors can be estimated and compensated for by combination with Conventional Monte Carlo (sections VII.F and VII.J). The other sections present miscellaneous ideas for computational improvements.

This section is intended to provide guidance for future work and show that solutions exist for future problems that may be encountered. Its purpose is not to stake any academic claims to these ideas.

VII.A. Parameter Screening: Identification of Statistically Insignificant Parameters

This section describes a simple procedure for identifying which of the uncertain parameters are significant and which are not. Insignificant parameters may be ignored.

The number of parameters is an important factor in the performance of the boundary-approximation procedure, and should be kept to a minimum. Unfortunately, the number of parameters increases roughly as the square of the number of components: The feed flow uncertainties require C+1 parameters. Tray efficiency adds another, and the operating uncertainties add several more, e.g., flooding rate, column pressure drop, heat exchanger heat transfer coefficients. The thermodynamic database solution model is the main culprit—the Modified Wilson (Tsuboka and Katayama, 1975) solution model requires C(C-1) parameters. Thus, for a 10-component system, there are more than 100 parameters. The current procedure cannot handle such large dimensions.

However, as Kubic and Stein (1988) state: “Not all uncertainties are created equal.” For example, many of those 10 components should distribute in the column with no effect on the key component separations. It follows that the feed flowrate and database parameters for the key components will be significant, while the other component parameters will be marginally significant or insignificant. Insignificant parameters can be ignored without major loss in accuracy.

A parameter ‘screening’ procedure is required to separate the significant parameters from the insignificant ones. The following approach requires little initial computational effort.

Table VII.A-1
Parameter Screening Algorithm

(1) Temporarily ignoring the database uncertainties, perform the initialization procedure using only the feed flowrate and tray efficiency uncertainties.[1]

(2) Perform a Monte Carlo integration to calculate a base reliability estimate. Next, toggle each component flowrate uncertainty off and recalculate the reliability. If there is no change, this component is statistically insignificant.[2]

- We now have a list of the significant feed components. Tray efficiency is almost certainly significant.

(3) Determine which pairs of database parameters are significant (from the list of significant feed components only). Two approaches might be followed: (1) Check the principal axis of each pair's confidence interval at some maximum statistical distance in both directions. The pair is insignificant if the boundary is not encountered. However, this is risky because the boundary geometry may be more complicated (statistically significant in other directions). A safer, but more time-consuming approach is to calculate the partial derivatives with respect to the binary parameters at all known boundary points. Then, refit the boundary-approximation including the additional parameter pair and recalculate the design reliability to see if it is changed significantly.

VII.B. Approximate Handling of Marginally Significant Parameters

This section suggests a modification to the boundary-approximation procedure that allows marginally significant parameters to be handled using linear extrapolation, without incurring major dimensional increase.

Statistically insignificant parameters may be ignored. However, marginally significant parameters should not be ignored if a very accurate reliability estimate is desired. A conservative approach is to set these parameters to their worst-case values and ignore them. A more accurate approach is to remove these parameters from the boundary search and geometric fit (thus avoiding memory doubling and slower convergence), but calculate the

[1] Include additional operating uncertainties such as flooding rate and pressure drop.

[2] Note: A cautious designer might run the convergence procedure for a few additional boundary points in the feed flowrate and tray efficiency space before toggling each component uncertainty. This would provide a more accurate estimate of the component significance. Also, it is possible for a significant component to have a positive effect in one location and a canceling negative effect in another. This can be checked for by looking at individual neighbor groups rather than just the overall reliability.

parameters' partial derivatives at the boundary points to provide a linear approximation. (Note: The base value could be either the design point θ^{Des} or a median value to provide the most accurate linear approximation.) The linear approximation could be varied throughout the constraint boundary region, perhaps using an averaged partial derivative value from the boundary points in each nearest-neighbor group. Finally, for the Monte Carlo integration, simulate random values for the marginal parameters then make a linear correction to the boundary-approximation residual (Eqs III.B.2.a-1 and III.B.2.a-3) by the partial derivative of the parameter(s) times the distance between the random point(s) and the base point.

VII.C. Improved Handling of the Thermodynamic Database Parameters

This section suggests alternative geometries for the thermodynamic parameters. This should decrease dimension and improve accuracy and convergence.

It was a decision of this research *not* to modify the geometry of the Modified Wilson (Tsuboka and Katayama, 1975) database parameter space, even though there are good reasons for doing so—namely that it is difficult to geometrically model such a highly correlated parameter space. Section V.C.1.b. observed that the database parameters may increase reliability error. Perhaps better geometries can avoid this.

VII.C.1. Database Parameter Transformation to Improve the Constraint Boundary Shape

When database parameters were included in the test problems, the percentage of nearest-neighbor groups with saddle curvature jumped to almost 100%. Saddle curvature is the most difficult to model accurately. Therefore if the geometry of the constraint boundary

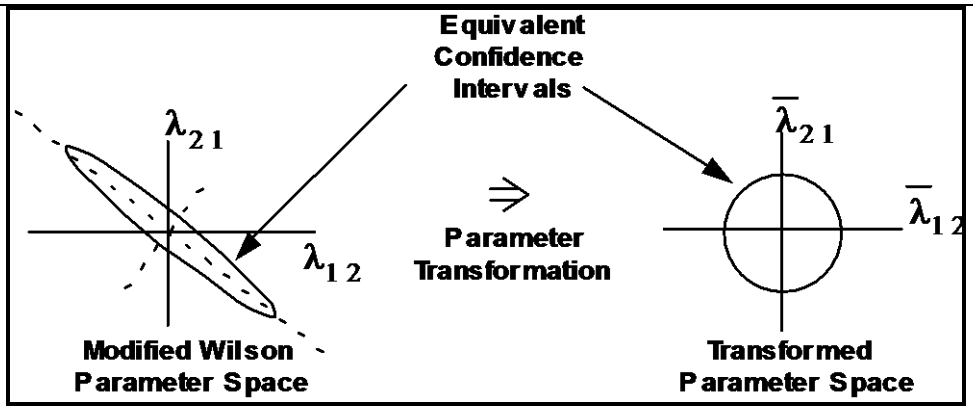


Figure VII.C.1-1
Parameter Transformation to 'Regularize' the Boundary Shape and Improve the Boundary-Approximation

can be altered so as to improve the curvature and make it more consistently convex or concave, the convergence rate and accuracy should improve. One approach is database parameter transformation. Each binary pair of Modified Wilson coefficients is highly correlated with a confidence interval shaped like a diagonal ellipse. If a new coordinate system is defined along the principal major and minor axes of this diagonal ellipse, two new parameters can be defined as shown in Figure VII.C.1-1. By proper scaling, the diagonal elliptical confidence interval can be transformed into a spherical one. Hopefully, the constraint boundary will also become more spherical for an easier approximation.

VII.C.2. Assumption of Symmetry

The test problem constraint boundaries appear to be fairly symmetrical along the database principal axis (Figures V.B.3-2, V.B.3-3 and Appendix C). That is, the constraint boundary shape in the top-left quadrant is roughly equivalent to the constraint boundary in the bottom-right quadrant. So one could assume that the constraint boundary is symmetrical, then converge the boundary-approximation in just one of the two quadrants (one of the eight hyperquadrants for the ternary component case, etc.). This will significantly improve convergence, but at the cost of increased error. However, error may be acceptably small, especially when the database uncertainties are marginally significant. A safer approach would be to proceed with the full nonsymmetrical boundary-approximation, then periodically check the accuracy of the symmetry approximation and switch to symmetry for final convergence if accuracy is acceptable.

VII.C.3. Binary Parameter Pair Replacement by a Single Parameter

Perhaps the simplest approach is to replace each database binary parameter pair with a single uncertain parameter. This is reasonable since the joint probability distribution is so highly correlated that it appears as a one-dimensional straight line (Figure V.B.3-1). This reduces the database dimension from C(C-1) to C(C-1)/2.

VII.D. Special Handling of Uncertain Parameters With Minimum or Maximum Bounds

Parameters with minimum or maximum bounds do not naturally fit into radial coordinate systems. Cylindrical coordinate systems might be more efficient.

Certain uncertain parameters, such as tray efficiency, have minimum or maximum bounds. Special handling can avoid dimensional doubling and thus improve performance, see section V.B.2.a. The current procedure requires a patch, but better coding can minimize memory and handle this automatically.

More efficient performance might be obtained by switching to a cylindrical coordinate system. For example, the feed flowrate space would continue with radial coordinates, but tray efficiency would use rectangular coordinates. The next-point selection procedure would

continue as before in the feed flowrate space. But for each radial direction, we might instead search for *two* boundary points at the minimum and maximum tray efficiency values.

VII.E. Convergence Acceleration by Curve-Fit Extrapolation

By monitoring the converging reliability estimate, it might be possible to accelerate convergence by curve-fitting and extrapolating to the final value.

As observed in section V.C.2.a., the boundary-approximation procedure quickly corrects any gross initial error (>1.0%) in the reliability estimate, but may take significantly longer to eliminate the remaining error and converge. It might be possible to curve-fit the converging reliability estimate to some factor such as the number of boundary points, then accelerate convergence by extrapolating this curve to infinity. Development and testing should be performed with geometrical problems. See Appendix D-2 for the convergence plots with distillation test problems.

VII.F. Estimating Final Error in the Boundary-Approximation Reliability Estimate

If very high accuracy in the reliability estimate is critical, final error can be estimated and corrected for by performing a Conventional Monte Carlo run using a small random set and rechecking these points with the boundary-approximation procedure. It is possible to estimate a confidence interval for the error, however the statistical work is incomplete and the confidence intervals are high in comparison to the typical errors encountered.

A fundamental limitation of the boundary-approximation procedure is that final or systematic error—though small for all test problems, see section V.C.1.b.—is unknown. One would therefore like to check the accuracy of the reliability estimate. This can be done by performing a short Conventional Monte Carlo simulation then rechecking these points with the boundary-approximation procedure. Depending on the number of points checked and the number of errors found, a confidence interval for the final error can be estimated.

Confidence Interval When No Erroneous Points are Found

Even when no erroneous points are found, error may still exist. However, we can determine an upper bound for the error. Define $\mathfrak{R}_\Delta = \left| \mathfrak{R}^{True} - \mathfrak{R}^{BA} \right|$, which is the absolute value of the final error in the boundary-approximation procedure. For M Monte Carlo points examined, the probability of finding k errors is given by the binomial distribution:

$$\Pr\left\{ b\left(M, \mathfrak{R}_\Delta\right) = k \right\} \qquad \text{Eq VII.F-1}$$

When no errors occurred, the confidence interval for $\mathfrak{R}_\Delta$ is $(0, \mathfrak{R}_\Delta^{Max})$ as defined by:

$$\Pr\left\{ b\left(M, \mathfrak{R}_\Delta^{Max}\right) = 0 \right\} = \frac{\alpha}{2} \qquad \text{Eq VII.F-2}$$

$$\binom{M}{0}\left(\mathfrak{R}_{\Delta}^{Max}\right)^{0}\left(1-\mathfrak{R}_{\Delta}^{Max}\right)^{M-0} = \left(1-\mathfrak{R}_{\Delta}^{Max}\right) = \frac{\alpha}{2} \qquad \text{Eq VII.F-3}$$

$$\mathfrak{R}_{\Delta}^{Max} = 1 - exp\left\{ln\left(\frac{\alpha}{2}\right)\Big/M\right\} \qquad \text{Eq VII.F-4}$$

The confidence interval is $\alpha/2$ because this is actually a two-sided confidence interval complicated by the absolute value.

Some values from Eq VII.F-4 are shown in Table VII.F-1. Note that even if 100 Monte Carlo points are tested, the uncertainty is ±3.62%, which is large in comparison to the reliability errors that have been found. Therefore this approach is of little use unless a large number of Monte Carlo points is tested.

Table VII.F-1
95% Upper Bound for Final Error

M	Maximum Error (±%)
10	30.8
20	16.8
50	7.11
100	3.62
250	1.46
500	0.74
1,000	0.37
10,000	0.037

Confidence Interval When Erroneous Points are Found

When erroneous points are found, the confidence interval will be even wider. Statistically, this case is very complicated. For every Monte Carlo point checked, there are three outcomes: No error; positive error; negative error. For M points, we have a trinomial distribution:

$$t\{M, \mathfrak{R}_{+}, \mathfrak{R}_{-}\} = k_{+}, k_{-} \qquad \text{Eq VII.F-5}$$

where

$\mathfrak{R}_{+}$ is the probability of positive error (overestimation of reliability)

$\mathfrak{R}_{-}$ is the probability of negative error (underestimation of reliability)

k_{+} is the number of positive error Monte Carlo points

k_{-} is the number of negative error Monte Carlo points

and we wish to determine a confidence interval for the function:

$$\mathfrak{R}_{\Delta} = \mathfrak{R}_{+} - \mathfrak{R}_{-} \qquad \text{Eq VII.F-6}$$

where $\mathfrak{R}_{\Delta}$ is the *net* or final error in the reliability estimate

I found this calculation too difficult to solve rigorously, and it was not pursued. It is likely that confidence intervals will be too wide to be of much use.

Perhaps the most straightforward way to handle erroneous points is to fix the boundary-approximation where it is incorrect. This can be done by performing a line-search in the

direction of the erroneous point to get a new boundary point that improves the approximation in that region. This avoids the statistical difficulties.

VII.G. Avoiding Expensive Finite Difference Calculation of the Boundary Gradient by Using the Converged Jacobian Matrices

As dimension increases, the nested-loop line-search algorithm will become inefficient because the boundary point gradient requires p finite-difference simulations. An alternative is to calculate the gradient from the process model Jacobian matrices, some of which may already be available from the process solution model.

While the nested-loop line-search algorithm (section III.E.1) was satisfactory for this research, it becomes slower as dimension increases, because approximately p+6 converged process simulations are required per boundary point.[1] This might become a bottleneck for large dimensions. One solution is to use the simultaneous algorithm instead (section III.E.2). Another is to calculate the gradient information using the process model Jacobian matrices as provided by Eq III.E.2.a.i-17 and repeated below. This is computationally much faster, because the process model is not repeatedly converged, and some of the Jacobian matrices are already available from the iterative solution procedure (section IV.A.3).[2]

$$\text{GradientVector} = \frac{\partial g_i}{\partial \theta} - \frac{\partial g_i}{\partial x}\left(\frac{\partial \mathbf{h}}{\partial x}\right)^{-1}\frac{\partial \mathbf{h}}{\partial \theta} \qquad \text{Eq VII.G-1}$$

The Jacobian calculations are performed at the boundary point. $(\partial \mathbf{h}/\partial x)^{-1}$ is already available from the Newton iterative solution of the process model. $\partial g_i/\partial \theta$ and $\partial g_i/\partial x$ are analytically simple, with most components being zero. $\partial \mathbf{h}/\partial \theta$ is more complex, but is not difficult to calculate: Where θ are the feed flowrates, $\partial \mathbf{h}/\partial \theta$ is analytically simple. Where θ is the tray efficiency, $\partial \mathbf{h}/\partial \theta$ can be calculated analytically. Where θ are the database thermodynamic database parameters, finite-differencing may be necessary. However, these are done one distillation stage at a time, and are much faster than a single distillation iterative step.

---

1 Approximately 6 simulations are required to converge the line-search to the boundary point—at least 4 and perhaps 8-9 at most. p simulations are required to perform the gradient calculation using finite-differencing.

2 The Naphtali-Sandholm (1971) algorithm was used for distillation, which is a straightforward Newton iterative method and provides the Jacobian matrix $(\partial \mathbf{h}/\partial x)^{-1}$. Other methods, such as hybrid-Newton or Sum-Rates, do not automatically provide Jacobian matrices.

VII.H. Changing Inequality Constraints

If any of the process inequality constraints are changed (e.g., what is the reliability for 99% overhead purity, 98%, 95%, … ?) the boundary-approximation procedure must be repeated. Currently, this requires deleting all boundary points with that particular active constraint, then finding new boundary points and refitting the boundary-approximation. This could also occur within a larger flowsheet analysis, where one process operation is relaxed for another to take up the slack.

If the inequality constraint is only changed slightly, a linear extrapolation from the current boundary-approximation might be sufficiently accurate—at least for a first approximation sensitivity analysis. This may be handled without major code modification because the boundary point gradient vectors already provide this information.

Another approach is to define a 'slack' parameter and add it to the inequality constraint directly, i.e., similarly to Eq II.C.4-1. Then, converge the boundary-approximation with this parameter being varied over the appropriate range using a uniform distribution. Then, perform the Monte Carlo integration with this slack parameter fixed at single values. This will provide an estimate of the reliability for a range of inequality constraint specifications, e.g., from 95% to 99% overhead purity.

VII.I. Optimization of Convergence Criteria to Maintain Accuracy While Minimizing Computational Effort

The various convergence criteria were set at tightest tolerances for testing purposes because accuracy was paramount. Loosening these tolerances should provide a substantial increase in computational efficiency without major loss in accuracy.

For this research, all convergence criteria were set very tightly since accuracy was most important. However, this requires the greatest computational effort. Unfortunately, preliminary sensitivity studies showed that convergence and final accuracy deteriorated when the criteria were loosened. However, initial accuracy remained satisfactory, so if high accuracy is not critical, the convergence criteria can be loosened for significantly faster execution.

In the nested-loop line-search algorithm, the interactions between the convergence criteria of the process simulation model (inner loop) and line-search (outer loop) should be explored for further optimization. For example, the settings for the inner loop could be loosened for the first few checkpoints. However, the line-search algorithm would have to be modified to handle less accurate interpolation points.

Also, the convergence criteria tolerance for distillation in the line-search may be different from the tolerance in the gradient calculations. In particular, by increasing the finite-difference step length one can loosen the tolerance and still get accurate gradients with less effort, so long as second-order curvature is insignificant, see section IV.B.4.

Rigorous studies should be performed with geometrical test problems to determine the level of accuracy required in the point+gradient information.

VII.J. Improving Computational Efficiency of Conventional Monte Carlo

If Conventional Monte Carlo *is* used, it is possible to improve computational efficiency by: (1) sorting the random points before simulating, and (2) loosening the process simulation convergence criteria, both of which reduce the number of iterations required to converge each point. This section present some studies.

This research has shown that the boundary-approximation procedure is superior to Conventional Monte Carlo. However, Conventional Monte Carlo may still be useful in some situations: (1) It is simple and requires no development; (2) it serves as a check on the boundary-approximation procedure and allows estimation of any final error, see section VII.F; (3) a particular problem may have a very large number of parameters, for which the boundary-approximation procedure is still untested. If Conventional Monte Carlo is used, there are ways to make it run faster.

One option is to loosen the convergence criterion for the process simulation model, which allows each simulation point to converge in fewer iterations. This is less accurate since some points might be incorrect, but because errors occur in both directions (points that are actually inside the boundary converging outside, and vice-versa), they might cancel to leave a small net error.

This was tested with the Acetone-Benzene test problem (feed + tray efficiency + database uncertainties). Figure VII.J-1 shows the converging reliability estimates at different distillation convergence tolerances. The most accurate simulation is the heavy line (ε_{Dist}=2e-7, see Eq IV.A.4-1), which took 21.6 seconds per point. A second medium-accuracy simulation used ε_{Dist}=1e-2, and took 15.5 seconds per point. A third low-accuracy simulation used ε_{Dist}=1e0 (at this level, it converges after a single Newton step), and took 3.0 seconds per point. The medium accuracy simulation is only slightly in error (+1.2%), while the low accuracy simulation is very poor (+8.0%).

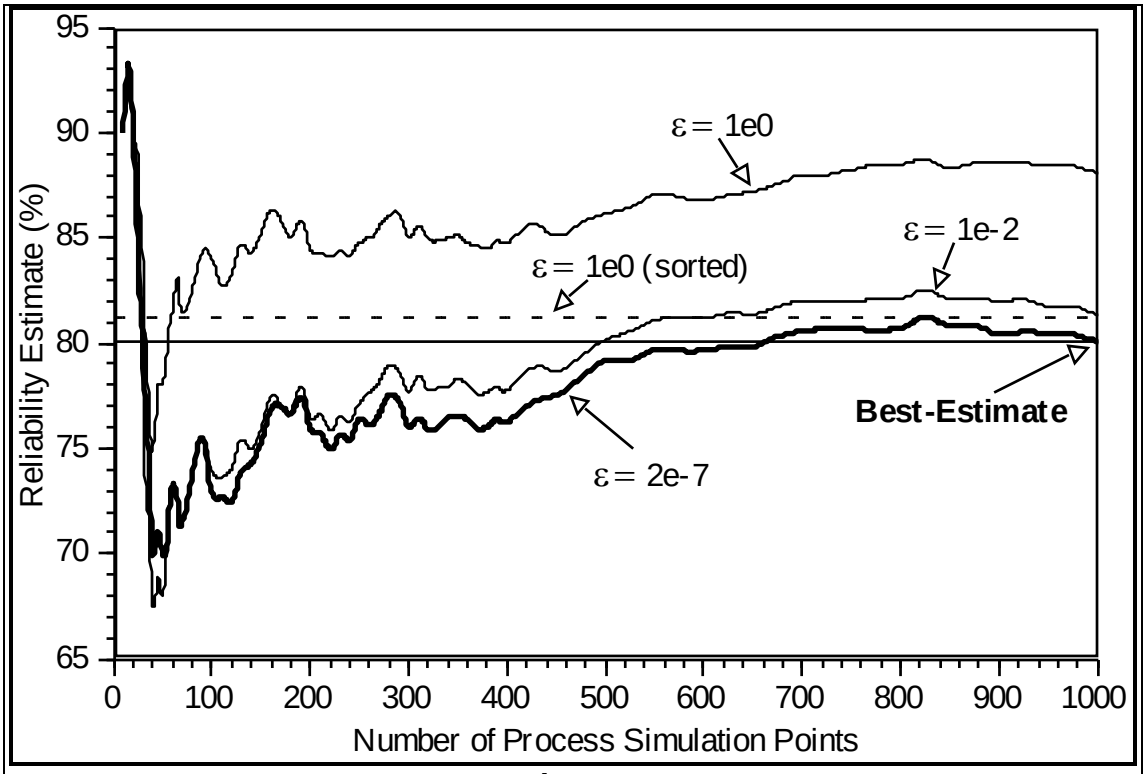


Figure VII.J-1
Inaccuracy in Conventional Monte Carlo as a Function of Distillation Convergence Criteria

As the distillation convergence tolerance is loosened, the reliability estimate becomes systematically high. This occurs because the distillation iteration does not oscillate around the true value. Rather, it proceeds in a constant 'direction' towards the true value without overshooting. The looser the convergence tolerance, the sooner the iteration stops and the further it is from the true point. Since reliability is usually greater than 50%, the iteration is more likely to start inside the constraint boundary and travel outside, rather than vice-versa.[1] So it is more likely to converge inside the boundary when the true point is outside than vice-versa, causing over-optimistic reliability estimates.

One partial fix is to sort the random points before simulating. This does two things: First, convergence is significantly faster, because the iteration does not have to step as far from one simulation point to the next—initial estimates are better. Second, the reliability estimate

---

[1] The column profile is not reset between simulation points — the previous simulation point's converged profile is used as the initial estimate for the current point. Another option is to reset the column profile to some base value, e.g., $\boldsymbol{\theta}^{Des}$, however this will not eliminate the systematic error.

should be more accurate, as the number of inside-to-outside and outside-to-inside iteration paths should be more equal.

This was checked with the Acetone-Benzene test problem and the loosest tolerance case (θ^{Des} =1e0), and is shown as the horizontal dashed line in Figure VII.J-1. The error was significantly reduced from +8.0% to +1.1%, which is excellent considering each point is only allowed one iteration step. And it was very fast, taking 2.67 seconds per point—eight times faster than the highest-accuracy simulation.

Sorting becomes difficult when there are many parameters. Some method of 'sub-sorting' is required (first sort using the most significant parameter, then sort smaller sections by the second significant parameter, etc.). However, this is cumbersome, and since the generating space is larger, iterations have to step farther between simulation points. So sorting will provide less of an improvement as dimension increases.

Alternatively, we could develop a smart procedure that converges each Monte Carlo point to a loose tolerance, then checks the inequality constraint. If it is far from zero, continue to the next point. But if the inequality constraint residual is near zero, converge to a tighter tolerance. Thus, obvious points are handled quickly and points close to the boundary are checked accurately.

An even smarter option is to calculate the inequality constraint after each iteration of the process model simulation, then track its progress and extrapolate to see if the inequality residual is approaching zero.

VIII. References

> If you copy something out of one book, it's plagiarism. If you copy it out of two, it's research. If you copy it out of three, you are a professor. Time Magazine, 1989 Dec 24.

Árva, P., and Csukás, B., 1987. "Fuzzy Valuated Relational Algebrae in Chemical Engineering". *Hungarian Journal of Industrial Chemistry*, 15:1-8.

Árva, P., and Czukás, B., 1988. "Computer-Aided Process Synthesis by Fuzzy Valuated Structures". *Computers & Chemical Engineering*, 12(2/3):215-224.

Barbosa, D., and Doherty, M. F., 1988a. "Design and Minimum-Reflux Calculations for Single-Feed Multicomponent Reactive Distillation Columns". *Chemical Engineering Science*, 43(7):1523.

Barbosa, D., and Doherty, M. F., 1988b. "Design and Minimum-Reflux Calculations for Double-Feed Multicomponent Reactive Distillation Columns". *Chemical Engineering Science*, 43(9):2377.

Biddulph. M. W., and Ashton, N., 1977. "Deducing Multicomponent Distillation Efficiencies from Industrial Data". *The Chemical Engineering Journal*, 14:7-15.

Biddulph, M. W., Kalbassi, M. A., and Dribika, M. M., 1988. "Multicomponent Efficiencies in Two Types of Distillation Column". *AIChE Journal*, 34(4):618-625.

Biegler, L. T., 1984. "Simultaneous Modular Simulation and Optimization". *Proceedings of the Second International Conference on Foundations of Computer-Aided Process Design*, (Westerberg and Chien, Eds). CACHE, 369-409.

Biegler, L. T., 1988. "On the Simultaneous Solution and Optimization of Large Scale Engineering Systems". *Computers & Chemical Engineering*, 12(5):357-369.

Bolles, W. L., and Fair, J. R., 1979. "Performance and Design of Packed Distillation Columns". *Third International Symposium on Distillation*, Apr 3-6, London.

Brignole, E. A., Gani, R., and Romagnoli, J. A., 1985. "A Simple Algorithm for Sensitivity and Operability Analysis of Separation Processes". *Industrial & Engineering Chemistry, Process Design and Development*, 24(1):42.

Broyden, C. G., 1965. "A Class of Methods for Solving Nonlinear Simultaneous Equations". *Mathematics of Computations*, 19:577-593.

Buzzi-Ferraris, G., and Tronconi, E., 1985. "Operational Optimization and Sensitivity Analysis of Multistage Separators". *Industrial & Engineering Chemistry, Process Design and Development*, 24(1):112.

Chan, H., and Fair, J. R., 1984a. "Prediction of Point Efficiencies on Sieve Trays. 1.Binary Systems". *Industrial & Engineering Chemistry, Process Design and Development*, 23:814-819.

Chan, H., and Fair, J. R., 1984b. “Prediction of Point Efficiencies on Sieve Trays. 2. Multicomponent Systems”. *Industrial & Engineering Chemistry, Process Design and Development*, 23:820-826.

Chen, H. -S., and Stadherr, M. A., 1985a. “A Simultaneous-Modular Approach to Process Flowsheeting and Optimization. Part I: Theory and Implementation”. *AIChE Journal*, 31(11):1843.

Chen, H. -S., Stadherr, M. A., 1985b. “A Simultaneous-Modular Approach to Process Flowsheeting and Optimization. Part II: Performance on Simulation Problems”. *AIChE Journal*, 31(11):1857.

Chen, H. -S., Stadherr, M. A., 1985c. “A Simultaneous-Modular Approach to Process Flowsheeting and Optimization. Part III: Performance on Optimization Problems”. *AIChE Journal*, 31(11):1868.

Chou, S. -M., and Yaws, C. L., 1988. “Minimum Reflux for Complex Distillations”. *Chemical Engineering*, 95(6):79.

Chu, J. C., Getty, R. J., Brennecke, L. F., and Paul, R., 1950. *Distillation Equilibrium Data*. New York. As found in Gmehling and Onken (1977ff).

Clopper, and Pearson, 1934. *Biometrika*, 26:404. (Referenced in Diem and Lentner, 1975, p186).

Conover, W. J., 1980. *Practical Nonparametric Statistics*, 2nd ed., John Wiley & Sons, New York.

Dankworth, D. C., and Sundaresan, S., 1989. “A Macroscopic Model for Countercurrent Gas-Liquid Flow in Packed Columns”. *AIChE Journal*, 35(8):1282.

Darby-Dowman, K., Lucas, C., Mitra, G., and Yadegar, J., 1988. “Linear, Integer, Separable and Fuzzy Programming Problems: A Unified Approach towards Reformulation”. *Journal of the Operational Research Society*, 39(2):161-171.

Deák, I., 1988. “Multidimensional Integration and Stochastic Programming”. *Numerical Techniques for Stochastic Optimization*, (Ermoliev, Yu, Wets, R. J-B Eds). Springer-Verlag, New York. 187-200.

Diaz, A., 1987. “Interactive Solution to Multiobjective Optimization Problems”. *International Journal for Numerical Methods in Engineering*, 24:1865-1877.

Diem, K., and Lentner, C., 1975. *Scientific Tables,* 7th Ed, Ciba-Geigy Limited, Basle, Switzerland.

Diwekar, U. M., and Rubin, E. S., 1991. “Stochastic Modeling of Chemical Processes”. *Computers & Chemical Engineering*, 15(2):105-114.

Dovi, V. G., and Paladino, O., 1990. “Use of Fuzzy Optimization Algorithms for the Solution of a Class of Ill-Posed Problems in Data Analysis”. *Computers & Chemical Engineering*, 14(9):957-966.

Dubois, D., and Prade, H., 1988. *Possibility Theory. An Approach to Computerized Processing of Uncertainty*, Plenum Press, New York.

Duran, M. A., and Grossmann, I. E., 1986. “Simultaneous Optimization and Heat Integration of Chemical Processes”. *AIChE Journal*, 32(1):123-138.

Eduljee, H. E., Rao, V. N. K., and Rao, M. N., 1958. *Chemical and Engineering Data Series*, 3:44. As found in Gmehling and Onken (1977ff).

Ellis, M. F., Koshy, R., Muares, G., Gomez-Munoz, A., and Holland, C. D., 1986. “Use of Multipoint Algorithm and Continuation Methods in the Solution of Distillation Problems”. *Computers & Chemical Engineering*, 10(5):433-443.

Fair, J. R., 1977. “Advances in Distillation System Design”. *Chemical Engineering Progress,* 73(11):78-83.

Fair, J. R., Null, H. R., and Bolles, W. L., 1982. “Scale-up of Plate Efficiency from Laboratory Oldershaw Data”. *Industrial & Engineering Chemistry, Process Design and Development*, 22:53-58.

Fedrizzi, M., 1987. “Introduction to Fuzzy Sets and Possibility Theory”. *Optimization Models Using Fuzzy Sets and Possibility Theory*, (Kacprzyk, J, and Orlovski, S, A., Eds). D. Reidel Publishing Company, Dordrecht, Holland. 13-26.

Fisher, W. R., Doherty, M. F., and Douglas, J. M., 1985a. “Effect of Overdesign on the Operability of Distillation Columns”. *Industrial & Engineering Chemistry, Process Design and Development*, 24(3):593.

Fisher, W. R., Doherty, M. F., and Douglas, J. M., 1985b. “Evaluating Significant Economic Trade-Offs for Process Design and Steady-State Control Optimization Problems”. *AIChE Journal*, 31(9):1538-1547.

Fisher, W. R., Doherty, M. F., and Douglas, J. M., 1988a. “The Interface Between Design and Control. 1. Process Controllability”. *Industrial & Engineering Chemistry Research*, 27(4):597-605.

Fisher, W. R., Doherty, M. F., and Douglas, J. M., 1988b. “The Interface Between Design and Control. 2. Process Operability”. *Industrial & Engineering Chemistry Research*, 27(4):606-611.

Fisher, W. R., Doherty, M. F., and Douglas, J. M., 1988c. “The Interface Between Design and Control. 3. Selecting a Set of Controlled Variables”. *Industrial & Engineering Chemistry Research*, 27(4):611-615.

Fitzmorris, R. E., and Mah, R. S. H., 1980. “Improving Distillation Column Design Using Thermodynamic Availability Analysis”. *AIChE Journal*, 26(2):265.

Floudas, C. A., 1987. “Simultaneous Process Synthesis and Control of Chemical Processes”. *1987 AIChE Annual Meeting*, New York, Paper 86B.

Floudas, C. A., Ciric, A. R., and Grossmann, I. E., 1986. “Automatic Synthesis of Optimum Exchanger Network Configurations”. *AIChE Journal*, 32(2):276.

Free, K. W., and Hutchison, H. P., 1959. *Journal of Chemical and Engineering Data*, 4:193. As found in Gmehling and Onken (1977ff).

Gaines, B. R., 1984. “Fundamentals of Decision: Probabilistic, Possibilistic and Other Forms of Uncertainty in Decision Analysis”. *Fuzzy Sets and Decision Analysis*, (Zimmermann, H. -J., Zadeh, L. A., Gaines, B. R., Eds). North-Holland, New York, 47-65.

Ganesh, N., and Biegler, L. T., 1987. “A Reduced Hessian Strategy for Sensitivity Analysis of Optimal Flowsheets”. *AIChE Journal*, 33(2):282-296.

Gmehling, J., and Onken, U., 1977ff. *Vapor-Liquid Equilibrium Data Collection*. DECHEMA Chemistry Data Series, Frankfurt, Germany.

Godsey, D. D., 1987. *The Study of the Effect of Vapor-Liquid Equilibrium Uncertainty on Distillation Tower Design Confidence and a Comparison of Conventional and Statistical Experimental Design*. MS Thesis, University of Kansas.

Goodman, I. R., and Nguyen, H. Y., 1985. *Uncertainty Models for Knowledge-Based Systems*, Elsevier Science Publishing Company, New York.

Grossmann, I. E., and Floudas, C. A., 1987. “Active Constraint Strategy for Flexibility Analysis in Chemical Processes”. *Computers & Chemical Engineering*, 11(6):675-693.

Grossmann, I. E., Halemane, K. P., and Swaney, R. E., 1983. “Optimization Strategies for Flexible Chemical Processes”. *Computers & Chemical Engineering*, 7(4):439-462.

Grossmann, I. E., and Morari, M., 1984. “Operability, Resiliency and Flexibility - Process Design Objectives for a Changing World”. *Proceedings of the Second International Conference on Foundations of Computer-Aided Process Design*, (Westerberg, A. W. and Chien, H. H., Eds). CACHE, 931-1010.

Grossmann, I. E., and Sargent, R. W. H., 1978. “Optimum Design of Chemical Plants with Uncertain Parameters”. *AIChE Journal*, 24(6):1021-1028.

Grottoli, M. G., Biardi, G., and Pellegrini, L., 1991. “A New Simulation Model for a Real Trays Absorption Column”. *Computers & Chemical Engineering*, 15(3):171-179.

Halemane, K. P., and Grossmann, I. E., 1983. “Optimal Process Design under Uncertainty”. *AIChE Journal*, 29(3):425-433.

Hankinson, R. W., and Thomson, G. H., 1979. “A New Correlation for Saturated Densities and their Mixtures”. *AIChE Journal*, 25(4):653-663.

Henley, E. J., and Seader, J. D., 1981. *Equilibrium-Stage Separation Operations in Chemical Engineering*. John Wiley & Sons, New York.

Hernandez, M. R., Gani, R., Romagnoli, J. A., and Brignole, E. A., 1984. “The Prediction of Properties and Its Influence on the Design and Modeling of Superfractionators”. *Proceedings of the Second International Conference on Foundations of Computer-Aided Process Design*, (Westerberg and Chien, Eds). CACHE, 709-740.

Hirose, Y., Takahashi, M., and Tachibana, H., 1986. “Extension of Kubicek Algorithm and its Application to Distillation Problems”. *Computers & Chemical Engineering*, 10(1):1-5.

Hlavácek, V., and van Rompay, P., 1985. “Simulation of Countercurrent Separation Process via Global Approach”. *Computers & Chemical Engineering*, 9(4):343-350.

Hogg, R. V., and Craig, A. T., 1978. *Introduction to Mathematical Statistics,* 4th ed., Macmillan Publishing Co., Inc. New York.

Howat, C. S., 1983. *Phase Equilibria Studies of the Isoprene, n-Pentane, Cyclopentadiene and Dicyclopentadiene System from 290 to 360K*. Ph.D Thesis, University of Kansas.

Hufton, J. R., Bravo, J. L., and Fair, J. R., 1988. “Scale-Up of Laboratory Data for Distillation Columns Containing Corrugated Metal-Type Structured Packing”. *Industrial & Engineering Chemistry Research*, 27(11):2096-2100.

Jelínek, J., 1988. “The Calculation of Multistage Equilibrium Separation Problems with Various Specifications”. *Computers & Chemical Engineering*, 12(2/3):195-198.

Julka, V., and Doherty, M. F., 1988. “Design and Analysis of Multicomponent Distillation Columns - Minimum Reflux Calculations”. *1988 AIChE Annual Meeting*, Washington, DC, Paper 78B.

Kalos, M. H., and Whitlock, P. A., 1986. *Monte Carlo Methods*, John Wiley & Sons, New York.

Kastánek, F., and Standart, G., 1967. “Studies on Distillation: XX.* Efficiency of Selected Types of Large Distillation Trays at Total Reflux”. *Separation Science*, 2(4):439-486.

Kato, M., Sato, T., Konishi, H., and Hirata, M., 1971. *Journal of Chemical Engineering of Japan*, 4:311. As found in Gmehling and Onken (1977ff).

Kaufmann, A., and Gupta, M. M., 1985. *Introduction to Fuzzy Arithmetic. Theory and Applications*, Van Nostrand Reinhold Co., New York.

Kister, H. Z., and Gill, D. R., 1991. “Predict Flood Point and Pressure Drop for Modern Random Packings”. *Chemical Engineering Progress*, 87(2):32-42.

Kocis, G. R., and Grossmann, I. E., 1987. “Relaxation Strategy for the Structural Optimization of Process Flow Sheets”. *Industrial & Engineering Chemistry Research*, 26(9):1869-1880.

Kojima, K., Tochigi, K., Seki, H., and Watase, K., 1968. *Kagaku Kogaku*, 32:149. As found in Gmehling and Onken (1977ff).

Kraus, J., and Linek, J., 1971. *Collection of Czechoslovak Chemical Communications*, 36:2547. As found in Gmehling and Onken (1977ff).

Krishna, R., Nanoti, S. M., and Goswami, A. N., 1989. “Mass-Transfer Efficiency of Sieve Tray Extraction Columns”. *Industrial & Engineering Chemistry Research*, 28(5):642-644.

Krishna, R., and Standart, G. L., 1979. “Mass and Energy Transfer in Multicomponent Systems”. *Chemical Engineering Communications*, 3:201.

Krishnamurthy, R., and Taylor, R., 1985a. “Nonequilibrium Stage Model of Multicomponent Separation Processes. Part I. Model Description and Method of Solution”. *AIChE Journal*, 31(3):449-456.

Krishnamurthy, R., and Taylor, R., 1985b. “Nonequilibrium Stage Model of Multicomponent Separation Processes. Part II. Comparison with Experiment”. *AIChE Journal*, 31(3):456-465.

Krishnamurthy, R., and Taylor, R., 1985c. “Nonequilibrium Stage Model of Multicomponent Separation Processes. Part III. The Influence of Unequal Component–Efficiencies in Process Design Problems”. *AIChE Journal*, 31(12):1973-1985.

Kryuchkov, Y. V., Anisimov, V. Y., and Borisov, E. V., 1986. “Determining the Parameters of Complex Systems with Fuzzy Initial Information”. *Radioelektronika*, 29(5):18-22.

Kubic Jr., W. L., 1986. *The Effect of Uncertainties in Physical Properties on Chemical Process Design*, Ph.D. Dissertation, Lehigh University.

Kubic Jr., W. L., 1989. Personal Communication.

Kubic Jr., W. L., and Stein, F. P., 1986. “The Applications of Fuzzy Set Theory to Uncertainty in Physical Property Models”. *Fluid Phase Equilibria*, 30:111-118.

Kubic Jr., W. L., and Stein, F. P., 1987. “Concept of System Similarity in Chemical Engineering”. *AIChE Journal*, 33(12):1986-1992.

Kubic Jr., W. L., and Stein, F. P., 1988. “A Theory of Design Reliability and Fuzzy Sets”. *AIChE Journal*, 34(4):583-601.

Kumar, A., and Lucia, A., 1987. “Separation Process Optimization Calculations”. *Applied Numerical Mathematics*, 3:409-425.

Kuncíˇr, M., and Madron, F., 1990. “Identification and Verification of Distillation Mathematical Models by Operational Data Analysis”. *CHISA 90*, Prague, Aug 26-30.

Lashmet, P. K., and Szczepanski, S. Z., 1974. “Efficiency Uncertainty and Distillation Column Overdesign Factors”. *Industrial & Engineering Chemistry, Process Design and Development*, 13:103-106.

Levien, K. L., and Morari, M., 1987. “Internal Model Control of Coupled Distillation Columns”. *AIChE Journal*, 33(1):83.

Linnhoff, B., Dunford, B. H., and Smith, R., 1983. “Heat Integration of Distillation Columns into Overall Processes”. *Chemical Engineering Science*, 38(8):1175.

Lockett, M. J., 1986. *Distillation Tray Fundamentals*, Cambridge University Press, Cambridge.

Lucia, A., 1985. “Partial Molar Excess Properties, Null Spaces, and a New Update for the Hybrid Method of Chemical Process Design”. *AIChE Journal*, 31(4):558-566.

Lucia, A., and Macchietto, S., 1983. “New Approach to Approximation of Quantities Involving Physical Properties Derivatives in Equation-Oriented Process Design”. *AIChE Journal*, 29(5):705-712.

Lucia, A., Miller, D. C., and Kumar, A., 1985. “Thermodynamically Consistent Quasi-Newton Formulae”. *AIChE Journal*, 31(8):1381-1388.

Lucia, A., and Westman, K. R., 1984. “Low Cost Solutions to Multistage, Multicomponent Separation Problems by a Hybrid Fixed Point Algorithm”. *Proceedings of the Second International Conference on Foundations of Computer-Aided Process Design*, (Westerberg, A. W. and Chien, H. H., Eds). CACHE, 741-763.

Lucia, A., and Xu, 1988. “Solving Large Chemical Process Optimization Problems”. *1988 AIChE Annual Meeting*, Washington, DC, Nov 27- Dec 2, Paper 82A.

Macchietto, S., Maduabueke, G. I., and Szczepanski, R., 1988. “Efficient Implementation of VLE Procedures in Equation-Oriented Simulators”. *AIChE Journal*, 34(6):955-963.

Madron, F., 1987. “Statistical Analysis of Industrial Distillation Columns Operational Data - Measurement of Tray Efficiencies”. *CHISA '87*. Paper #1184.

McDaniel, R., Bassyoni, A. A., and Holland, C. D., 1970. “Use of the Results of Field Tests in the Modeling of Packed Distillation Columns and Packed Absorbers - III”. *Chemical Engineering Science*, 25:633-651.

McDonald, K. A., and Palazoglu, A. N., 1987. “Robustness Analysis of High Purity Distillation Control Using Highly Structured Correlated Model Uncertainty Descriptions”. *1987 AIChE Annual Meeting*, New York, Nov 15-20.

Medina, A. G., Ashton, N., and McDermott, C., 1978. “Murphree and Vaporization Efficiencies in Multicomponent Distillation”. *Chemical Engineering Science*, 33:331-339.

Medina, A. G., McDermott, C., and Ashton, N., 1979a. “Prediction of Multicomponent Distillation Efficiencies”. *Chemical Engineering Science*, 34:861-866.

Medina, A. G., Ashton, N., and McDermott, C., 1979b. “Hausen and Murphree Efficiencies in Binary and Multicomponent Distillation”. *Chemical Engineering Science*, 34:1105-1112.

Meerboer, B., 1969. “Recommended Test Mixtures for Distillation Columns”, *Institute of Chemical Engineers*, London. As found in Gmehling and Onken (1977ff).

Michelsen, M. L., and Mollerup, J., 1986. “Partial Derivatives of Thermodynamic Properties”. *AIChE Journal*, 32(8):1389-1392.

Miller, D. C., and Lucia, A., 1985. “The Behavior of Hybrid Fixed-Point Methods in a Chemical Process Design Environment”. *AIChE Journal*, 31(2):329-332.

Miller, D. G., 1964. “Estimating Vapor Pressure. A Comparison of Equations”. *Industrial and Engineering Chemistry*, 56(3):46-57.

Mueller, E., and Stage, H., 1946. *Journal. American Oil Chemists’ Society*, 23:177. As found in Gmehling and Onken (1977ff).

Moore, C., 1985. “A Reliable Distillation Column Control Analysis Procedure For Use During Initial Column Design”. *1985 AIChE Annual Meeting*, Nov 10-15.

Naphtali, L. M., and Sandholm, D. P., 1971. “Multicomponent Separation Calculations by Linearization”. *AIChE Journal*, 17(1):148-153.

Nishida, N., Ichikawa, A., and Tazaki, E., 1974. “Synthesis of Optimal Process Systems with Uncertainty”. *Industrial & Engineering Chemistry, Process Design and Development*, 13:209-214.

Nishida, N., and Ohtake, Y., 1986. “A Method for the Solution of the Bicriterion Mixed Integer Programming Problems”. *Computers & Chemical Engineering*, 10(2):119.

Othmer, D. F., Chudgar, M. M., and Levy, S. L., 1952. *Industrial and Engineering Chemistry*, 44:1872.

Palazoglu, A., and Arkun, Y., 1986. “A Multiobjective Approach to Design Chemical Plants with Robust Dynamic Operability Characteristics”. *Computers & Chemical Engineering*, 10(6):567.

Palazoglu, A., and Arkun, Y., 1987. “Design of Chemical Plants with Multiregime Capabilities and Robust Dynamic Operability Characteristics”. *Computers & Chemical Engineering*, 11(3):205-216.

Papoulias, S. A., and Grossmann, I. E., 1983. “A Structural Optimization Approach in Process Synthesis - I. Utility Systems”. *Computers & Chemical Engineering*, 7(6):695.

Palazoglu, A., Manousiouthakis, B., and Arkun, Y., 1985. “Design of Chemical Plants with Improved Dynamic Operability in an Environment of Uncertainty”. *Industrial & Engineering Chemistry, Process Design and Development*, 24(3):802.

Pistikopoulos, E. N., and Mazzuchi, T. A., 1990. “A Novel Flexibility Analysis Approach For Processes With Stochastic Parameters”. *Computers & Chemical Engineering*, 14(9):991-1000.

Pham, H. N., Ryan, P. J., and Doherty, M. F., 1989. “Design and Minimum Reflux for Heterogeneous Azeotropic Distillation Columns”. *AIChE Journal*, 35(10):1585.

Prado, M., and Fair, J. R., 1990. “Fundamental Model for the Prediction of Sieve Tray Efficiency”. *Industrial and Engineering Research*, 29(6):1031-1042.

Rinard, I. H., 1987. “Sensitivity to Modeling Errors in Steady-State Process Simulation”. *Computers & Chemical Engineering*, 11(6):707-712.

Rippin, D. W. T., 1992. “The Future of Process Operations”. *Computing and Systems Technology Division Communications (CAST)*, 15(1):9-13.

Robbins, L. A., 1991. “Improve Pressure-Drop Prediction with a New Correlation”. *Chemical Engineering Progress*, 87(5):87.

Rocha, J. A., Cárdenas, J. C., Sosa, C., and Rosales, J., 1991. “Preliminary Design of Sieve Tray Extraction Columns. 2. Determination of the Column Height . Overall Efficiency of Sieve Tray Extractors”. *Industrial & Engineering Chemistry Research*, 28(12):1879-1883.

Rollett, A. P., Elkaim, G., Toledano, P., and Senez, M., 1956. *Comptus Rendus*, 242:2560. As found in Gmehling and Onken (1977ff).

Rosanoff, M. A., Bacon, C. W., and White, R. H., 1914. *Journal. American Chemical Society*, 36:1803. As found in Gmehling and Onken (1977ff).

Rubinstein, R. Y., 1981. *Simulation and the Monte Carlo Method*, John-Wiley & Sons, New York.

Sealey, C. J., 1970. “The Optimal Design of a Laboratory Distillation Column for Efficiency Studies at Finite Reflux”. *Chemical Engineering Science*, 25:561.

Shanker, G., Howat III, C. S., Torres, R. R., and Swift, G. W., 1981. “Vapor-Liquid Equilibria for 2-Methylbutene-1 and 2-Methyl-1,3-Butadiene at 310.93, 316.48 and 322.03 K”. *Fluid Phase Equilibria*, 5:305-321.

Shenoi, S., Melton, A., and Fan, L. T., 1988. “Chemical Process Design in the Face of Incomplete Information”. *1988 AIChE Annual Meeting*, Washington, DC, Nov 27- Dec 2.

Sivasubramanian, M. S., Taylor, R., and Krishnamurthy, R., 1987. “A Nonequilibrium Stage Model of Multicomponent Separation Processes. Part IV. A Novel Approach to Packed Column Design”. *AIChE Journal*, 33(2):325-327.

Sivasubramanian, M. S., and Boston, J. F., 1988. “Rate-Based Separation Process Modeling Techniques”. *1988 AIChE Annual Meeting*, Washington, DC, Nov 27- Dec 2, Paper 82D.

Skogestad, S., and Morari, M., 1987a. “Implications of Large RGA Elements on Control Performance”. *Industrial and Engineering Research*, 26(11):2323-2330.

Skogestad, S., and Morari, M., 1987b. “Control Configuration Selection for Distillation Columns”. *AIChE Journal*, 33(10):1620.

Skogestad, S., and Morari, M., 1987c. “Design of Resilient Processing Plants - IX. Effect of Model Uncertainty on Dynamic Resilience”. *Chemical Engineering Science*, 42(7):765.

Skogestad, S., Lundström, P., and Jacobsen, E. W., 1990. “Selecting the Best Distillation Control Configuration”. *AIChE Journal,* 36(5):753.

Sridhar, L. N., and Lucia, A., 1989. “Tearing Algorithms for Separation Process Simulation”. *Computers & Chemical Engineering*, 14(8):901-905.

Straub, D. A., and Grossmann, I. E., 1990. “Integrated Stochastic Metric of Flexibility for Systems with Discrete State and Continuous Parameter Uncertainties”. *Computers & Chemical Engineering*, 14(9):967-985.

Streich, M., and Kistenmacher, H., 1979. “Property Inaccuracies Influence Low Temperature Designs”. *Hydrocarbon Processing*, 58(5):237.

Swaney, 1988. “Minimum Flows Analysis of Multicomponent Countercurrent Contacting”. *1988 AIChE Annual Meeting*, Washington, DC, Nov 27- Dec 2, Paper 78C.

Swaney, R. E., and Grossmann, I. E., 1985a. “An Index for Operational Flexibility in Chemical Process Design: Part I: Formulation and Theory”. *AIChE Journal*, 31(4):621-630.

Swaney, R. E., and Grossmann, I. E., 1985b. “An Index for Operational Flexibility in Chemical Process Design: Part II: Computational Algorithms”. *AIChE Journal*, 31(4):631-641.

Swartz, C. L. E., and Stewart, W. E., 1987. “Finite-Element Steady State Simulation of Multiphase Distillation”. *AIChE Journal*, 33(12):1977.

SYNDES, Chemical Process Simulator. 1987. Colin S. Howat Software V1.01. User’s Manual.

Tanaka, H., Uejima, S., and Asai, K., 1982. “Linear Regression Analysis with Fuzzy Model”. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-12(6):903-907.

Toor, H. L., and Burchard, J. K., 1960. “Plate Efficiencies in Multicomponent Distillation”. *AIChE Journal*, 6(2):202.

Tsuboka, T., and Katayama, T., 1975. “Modified Wilson Equation for Vapor-Liquid and Liquid-Liquid Equilibria”. *Journal of Chemical Engineering of Japan*, 8(3):181-187.

Turunen, I., Nyberg, T., Järveläinen, J., Linko, Y. -Y., Linko, P., and Dohnal, M.,1985. “Fuzzy Modeling in Biotechnology. Sucrose Inversion”. *The Chemical Engineering Journal*, 30:B51-B60.

Venkataraman, S., and Lucia, A., 1986. “Exploiting the Gibbs-Duhem Equation in Separation Calculations”. *AIChE Journal*, 32(7):1057-1066.

Venkataraman, S., and Lucia, A., 1987. “Avoiding Variable-Scaling Problems using the Gibbs-Helmholtz Equation”. *Computers & Chemical Engineering*, 11(1):73-76.

Venkataraman, S., and Lucia, A., 1988. “Solving Distillation Problems by Newton-Like Methods”. *Computers & Chemical Engineering*, 12(1):55-69.

Verhoeye, L., and De Schepper, H., 1973. *Journal of Applied Chemistry and Biotechnology*, 23:607. As found in Gmehling and Onken (1977ff).

Vickery, D. J., and Taylor, R., 1986. “Path-Following Approaches to the Solution of Multicomponent, Multistage Separation Process Problems”. *AIChE Journal*, 32(4):547-556.

Vickery, D. J., Ferrari, J. J., and Taylor, R., 1988. “An ‘Efficient’ Continuation Method for the Solution of Difficult Equilibrium Stage Separation Process Problems”. *Computers & Chemical Engineering*, 12(1):99-103.

Vital, T. J., Grossel, S. S., and Olsen, P. I., 1984a. “Estimating Separation Efficiency, Part One”. *Hydrocarbon Processing*, Oct:55-56.

Vital, T. J., Grossel, S. S., and Olsen, P. I.,1984b. “Estimating Separation Efficiency, Part Two”. *Hydrocarbon Processing*, Nov:147-153.

Vital, T. J., Grossel, S. S., and Olsen, P. I., 1984c. “Estimating Separation Efficiency, Part Three - Packed Columns”. *Hydrocarbon Processing*, Dec:75-78.

von Neumann, J., Kent, R. H., Bellinson, H. R., and Hart, B. I., 1941. “The Mean Square Successive Difference”. *The Annals of Mathematical Statistics*, 12:153-162.

Wachter, J. A., and Ko, T. K. T., 1988. “Minimum Reflux Behavior of Complex Distillation Columns”. *AIChE Journal*, 34(7):1164-1184.

Wait, R., and Landauro, J., 1988. “Parallel Algorithms for Multicomponent Separation Calculations”. *AIChE Journal*, 34(6):964-968.

Watson, D. F., 1988. “Natural Neighbor Sorting on the n-Dimensional Sphere”. *Pattern Recognition*, 21(1):63-67.

Wayburn, T. L. and Seader, J. D., 1984. “Solution of Systems of Interlinked Distillation Columns by Differential Homotopy-Continuation Methods”. *Proceedings of the Second International Conference on Foundations of Computer-Aided Process Design,* (Westerberg, A. W. and Chien, H. H., Eds). CACHE, 765-862.

Wen, C. Y., and Chang, T. M., 1968. “Optimal Design of Systems Involving Parameter Uncertainty”. *Industrial & Engineering Chemistry, Process Design and Development*, 7(1):49-53.

Westman, K. R., Lucia, A., and Miller, D. C., 1984. “Flash and Distillation Calculations by a Newton-Like Method”. *Computers and Chemical Engineering*, 8(3/4):219-228.

Yager, R. R., 1982. “Fuzzy Prediction based on Regression Models”. *Information Sciences*, 26(1):45-63.

Yang, D. R., Waller, K. V., Seborg, D. E., and Mellichamp, D. A., 1990. “Dynamic Structural Transformations for Distillation Control Configurations”. *AIChE Journal*, 36(9):1391.

Zadeh, L. A., 1968. “Probability Measures of Fuzzy Events”. *Journal of Mathematical Analysis and Applications*, 23:421-427.

Zadeh, L. A., 1978. “Fuzzy Sets as a Basis for a Theory of Possibility”. *Fuzzy Sets and Systems*, 1:3-28.

Zadeh. L. A., 1988. “Fuzzy Logic”. *Computer*, April, 83-93.

Zuiderwieg, F. J., 1982. “Sieve Trays. A View on the State of the Art”. *Chemical Engineering Science*, 37(10):1441-1464.

Appendix A. Fuzzy Statistics Literature Search

This section summarizes my observations while searching the fuzzy theory literature. Also, a list is provided of some chemical engineering articles using fuzzy theory.

The KU library catalogue was searched by topic. By far the most useful book was "Possibility Theory", by Dubois and Prade (1988). The first two chapters provide the best introduction to the theory that I have found. Also, Kaufmann and Gupta (1985) and Fedrizzi (1987) are good introductions.

Unfortunately, the literature jumps from introductory articles to inaccessible abstract mathematics. Since the theory is still very young and deals with so complex a subject as uncertainty, it has not been completely accepted. Thus, there are many philosophical and statistical argument articles dealing with the concept of fuzziness. However, they either make general background reading or are several orders of technical magnitude out of reach of this fuzzy neophyte, e.g., the "theory of uncertainty" books such as Goodman and Nguyen (1985).

Many promising article titles led to some hitherto unknown corners of the KU library. Unfortunately, they were all dead ends. In the final outcome, I relied almost solely on Dubois and Prade (1988) to grasp the theory, with my confidence buoyed by the miscellaneous nontechnical introduction articles.

Some of the main research areas are linguistics, imprecise logic, and expert systems with imprecise information. The closest related topics and most advanced 'number crunching' in terms of mathematical computations and statistics are in management decision making and optimization, cybernetics, and statistical analysis in the 'soft sciences', all of which have developed "fuzzy regression" and fuzzy optimization methods.

Kubic and Stein (1988) deserve praise for making a conceptual leap in the application of these theories to chemical engineering. I could not find anything in the literature even remotely related to their development. Their most innovative idea was the separate and parallel estimation of the possibility of design success *and* failure. This is an intuitive leap in comparison to the fuzzy theory literature, which generally seems to concentrate on the success rate or power of a conclusion, and is lacking in checks on the negative side. This parallels the general lack of attention usually devoted to type II error estimation in the traditional statistical literature.

Only five articles were found that applied fuzzy theory to engineering practice: (1) Turunen et al. (1985) made an interesting use of fuzzy statistics to extrapolate a very uncertain biological model with inadequate data. They used an expert computer system to

build a prediction model that combined an empirical equation regressed from actual data with their own qualitative opinions on what should occur in the extrapolation area.

(2) Àrva and Csukás (1988) applied fuzzy theory to process synthesis to reduce the "combinatorial explosion" problem, i.e., the large number of possible designs that result when many different alternatives exist.

Several of the East European countries are using fuzzy theory in chemical engineering. Unfortunately, all the articles are in Czech.

(3) Kryuchkov et al. (1986) used fuzzy statistics to maximize an objective function subject to inequality constraints. They used a Bayes statistics approach to show that the best design maximizes the certainty with which the inequalities are satisfied. This is equivalent to minimizing the possibility that the inequality is violated, which confirms the Kubic and Stein (1988) method.

(4) Dovi and Paladino (1991) applied a fuzzy constrained optimization procedure to the problem of matrix inversion and parameter estimation for "ill-posed" problems, e.g., singular matrices, highly correlated multiple parameters. The system is generally effective but computationally burdensome.

(5) Finally, Shenoi et al. (1988) presented a paper at the 1988 AIChE Annual Meeting entitled "Chemical Process Design in the Face of Incomplete Information." With such a promising title, and the mention of fuzzy statistics in the abstract, I had high hopes. Then I read it. Unfortunately, this article has nothing to do with chemical engineering beyond a weak analogy in the introduction comparing process design to "artistic composition – from a rough pencil sketch to the finished painting." How profound. This article deserves ridicule if only to avenge the poor unsuspecting listeners at the meeting who must have suffered so dreadfully during its presentation, trapped in their seats after realizing too late the true (lack of) content. This article is a classic case: Overblown misleading title to bring in the crowds (and look good on the resume), nice inoffensive introduction mentioning "chemical engineering" a few times, then hit them with long-winded, inappropriate, irrelevant, and technically incomprehensible information designed to thoroughly anesthetize. Finally, a quick example at the end to show the engineering 'relevance' (Hah!). This article is not chemical engineering at all, but a badly-ported relational database fuzzy application.

Appendix B. Multi-Dimensional Integration by Monte Carlo

This section describes how Monte Carlo integration may used to integrate a multidimensional probability distribution throughout a given region. Methods for estimating confidence intervals are presented and tested.

Assume that the constraint boundary is known. More specifically, a means must exist to determine if any particular point is inside or outside the constraint boundary. Also assume the probability distribution is known exactly.

The design reliability is equal to the integration of the statistical distribution throughout the interior of the constraint boundary:

$$\Re = \int_{\theta \in \mathbf{R}} \Pr\{\theta\} d\theta \qquad \text{Eq B-1a}$$

$$\int \ldots \int \Pr\{\theta_1, \ldots, \theta_p\} \; d\theta_1 \ldots d\theta_p \qquad \text{Eq B-1b}$$

where $\mathbf{R}$ is the region inside the constraint boundary

The function $\Pr\{\theta\}$ is frequently analytically nonintegrable. Numerical methods, such as gaussian quadrature, become inefficient for high dimensions (Kalos and Whitlock, 1986), and difficult for complex boundaries. Deák (1988) recommends Monte Carlo integration for problems with more than 5 dimensions. All other known methods are computationally prohibitive if reasonable accuracy is required.

Monte Carlo integration (Rubinstein, 1981; Kalos and Whitlock, 1986) is straightforward for any dimension and boundary complexity: Simply generate a set of random points according to the probability distribution function (pdf), and determine whether each point is inside or outside the boundary. The fraction of points inside the boundary is approximately equal to the pdf integrated over that region.

This section discusses computer development tests for the Monte Carlo integration. First, I tested the computer random number generator for randomness. Second, I compared several methods for estimating confidence intervals in the Monte Carlo integration and reliability estimate. Finally, a required calculation for comparing Conventional Monte Carlo and the boundary-approximation procedure is presented.

B.1. Random Number Generator — Tests for Randomness

This section checks the MacIntosh computer internal random number generator for randomness. The generator passes all tests examined.

If the random number generator is not random, the Monte Carlo integration could give erroneous results. Unfortunately, no computer generated sequence of numbers can be truly

random. However, a good generator will 'act' random, i.e., be "pseudorandom." The definition of a pseudorandom generator is that it passes all tests for randomness.

One simple test for randomness is the chi-square goodness-of-fit test (Hogg and Craig, 1978). First, divide the generating space of the random number into several regions. Then generate a set of random numbers and compare the number of points in each region with the expected probability for each region. This produces a chi-square statistic, which is used to test for nonrandomness (bias, nonuniformity) in the generator.

A run was performed with a uniform probability distribution generator in the range (0,1), with occurrences tabulated from 0 to 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99, 0.995, 0.999, to 1.0. This produces a $\chi^2(17)$ statistic. A total of 3.8×10^6 points was run, and the result is plotted in Figure B-1-1. The statistic starts out large and at one point violates the 95% confidence interval, which would lead to a rejection of the hypothesis of randomness. But as the number of points increases, the statistic decreases, leading us to accept the hypothesis of randomness for this test.

The "mean square successive difference test" is used to check for non-random trends or "drift" (Neumann et al., 1941, as found in Diem and Lentner, 1975. p193). For a normal distribution, the following statistic is calculated:

$$\frac{\sum_{i}^{M-1}\left(u_{i+1}-u_i\right)^2}{\sum_{i}^{M}\left(u_i-\bar{u}\right)^2} \qquad \text{Eq B-1-1}$$

where u is a set of random numbers and $\bar{u}$ is the mean of the set

Diem and Lentner (1975) provide confidence intervals for $M \leq 200$. The standard normal approximation can be used for $M > 200$. Violation of the upper bound confidence interval indicates nonrandom short-term cyclic factors. Violation of the lower bound confidence interval indicates nonrandom long-term factors.

A set of 10^6 numbers were generated and the Eq B-1-1 statistic plotted in Figure B-1-2. The first time this was done, the 95% confidence interval was violated for $500 < M < 1300$, however, the test was passed in the limit as $M \varnothing 10^6$. To determine whether the confidence

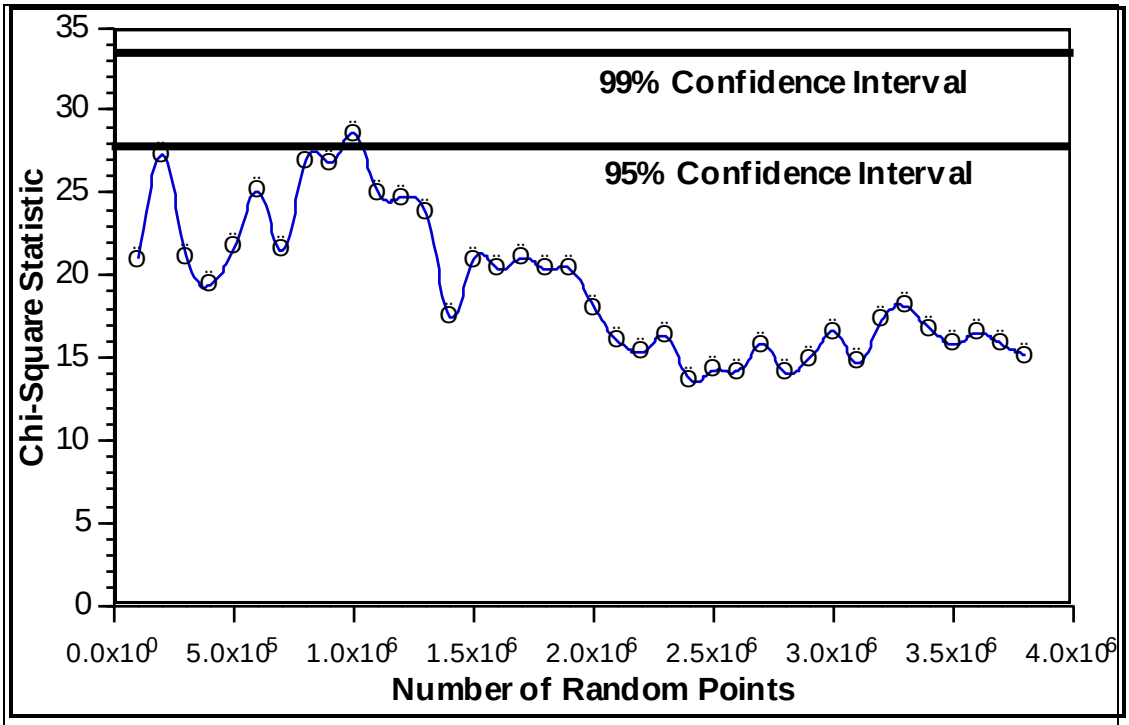


Figure B-1-1
Chi-Square Test for Randomness

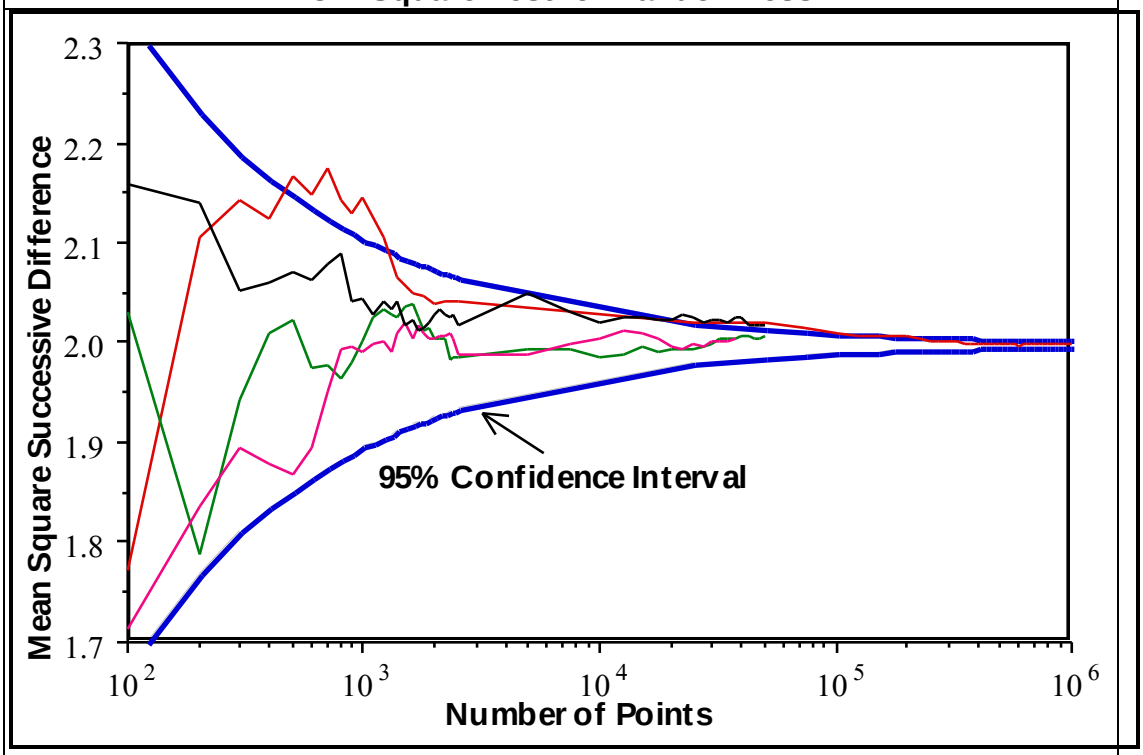


Figure B-1-2
Mean-Square-Successive-Difference Test for Randomness

interval violation was a 1 in 20 random phenomenon or a nonrandom effect, several other cases were run using different seed numbers. None of these cases violated the 95% confidence interval, as can be seen from the plot. Therefore, we may conclude that significant nonrandom trends do not exist.

The random number generator passed both the chi-square and mean square successive difference tests as $M \to \infty$. So even though the 95% confidence intervals were occasionally violated for low M, the generator appears to be satifactory.

Other statistical tests for randomness exist, see Rubinstein (1981). One simple, but powerful check is to generate pairs of numbers, then plot them as (*x*,*y*) coordinates on a computer screen. Nonrandomness shows up as uneven point scatter. The MacIntosh generator had no apparent problems.

B.2. Estimated Confidence Intervals of Reliability

This section presents and tests methods to estimate the confidence interval for the design reliability estimated by the Monte Carlo integration technique.

Each Monte Carlo random point produces a pass/fail result, depending on whether it is outside or inside the constraint boundary. The estimated reliability is the fraction of points inside the boundary. However, this fraction is an uncertain function of the actual reliability and the number of Monte Carlo points generated. Several methods exist to estimate this uncertainty.

The reliability estimate has a binomial distribution $b\{M,\Re\}$, where M is the number of Monte Carlo points, and $\Re$ is the design reliability. This has a discrete probability distribution function:

$$b\{M,\Re\} = \binom{M}{Y} \Re^{Y} (1-\Re)^{M-Y} \qquad \text{Eq B-2-1}$$

where Y is the number of successes, or points inside the boundary

$\frac{Y}{M}$ is an unbiased estimator of $\Re$ (Hogg and Craig, 1978).

The exact confidence interval is (Clopper and Pearson, 1934, as found in Diem and Lentner, 1975, p186):

$$Pr\{\Re_{Min} < \Re < \Re_{Max} \mid Y,M\} = 1-\alpha \qquad \text{Eq B-2-2}$$

$$\sum_{i=Y}^{M} \binom{M}{i} (\Re_{Min})^{i} (1-\Re_{Min})^{M-i} = \frac{\alpha}{2} \quad (\text{Solve for } \Re_{Min}) \qquad \text{Eq B-2-3a}$$

$$\sum_{i=0}^{Y}\binom{M}{i}\left(\Re_{Max}\right)^{i}\left(1-\Re_{Max}\right)^{M-i} = \frac{\alpha}{2} \qquad (\text{Solve for } \Re_{Max}) \qquad \text{Eq B-2-3b}$$

- α is level of significance or probability of Type I error (i.e., accepting a hypothesis that is incorrect). The designer must arbitrarily set this level. In this section, I use α =0.05 and α =0.01, for 95% and 99% confidence intervals, respectively.
- Eqs B-2-3a and b are solved iteratively for $\Re_{Min}$ and $\Re_{Max}$, respectively.

For computational purposes, the binomial pdf can be rewritten recursively (Diem and Lentner, 1975):

$$\Pr\{Y=0\} = \binom{M}{0}\Re^{0}(1-\Re)^{M-0} = (1-\Re)^{M} \qquad \text{Eq B-2-4a}$$

$$\Pr\{Y=i+1\} = \Pr\{Y=i\} \times \frac{\Re}{1-\Re} \times \frac{M-i}{i+1} \qquad \text{Eq B-2-4b}$$

Since $\Re$ is greater than 50% in practice, Y will be closer to M than to zero. So it is more efficient to perform the calculation starting from Y=M, and stepping downwards:

$$\Pr\{Y=M\} = \binom{M}{M}\Re^{M}(1-\Re)^{M-M} = \Re^{M} \qquad \text{Eq B-2-5a}$$

$$\Pr\{Y=i+1\} = \Pr\{Y=i\} \times \frac{1-\Re}{\Re} \times \frac{i}{M-i+1} \qquad \text{Eq B-2-5b}$$

For large M, the numbers get extremely small, resulting in computer underflow in the above calculation. This can be avoided by performing the calculations using logarithmic transformations and extended precision. But the computations are still very lengthy, especially when M is large, and ← is close to 50%.

To avoid the above iterative calculation, one simplification is to use Y/M as the estimate for $\Re$, then calculate the cumulative distribution function (cdf) by discrete summation of the pdf using Eq B-2-5ab. This method is inexact because Y/M is only an estimate for $\Re$. However, it becomes more accurate as M increases. Henceforth, it is referred to as Method 0.

For the special case where Y=M (all points pass), then the reliability estimate is 100%, the confidence interval is one-sided, and an exact confidence interval is possible. From Eq B-2-5a, we have:

$$\Pr\{Y=M\} = \left(\Re_{Min}\right)^{M} = \alpha \qquad \text{Eq B-2-6}$$

which can be solved exactly as:

$$\Re_{Min} = exp\left[ln\left(\frac{\alpha}{M}\right)\right] \quad (\text{when} \quad Y = M) \qquad \text{Eq B-2-7}$$

The above methods require major computer effort. An inner loop performs a logarithm, exponentiation, and summation inside a do-loop of order M. An outer loop iterates to solve for $\Re_{Min}$ and $\Re_{Max}$ (Method 0 avoids the outer loop.). For large M, the expense can be prohibitive.

Luckily, there are much cheaper, albeit approximate methods. One such method is the standard normalization technique. The mean and variance of the binomial distribution are $M\Re$ and $M\Re(1-\Re)$ respectively. Assuming the distribution is normal, a confidence interval for $\Re$ is:

$$\Re = \frac{Y}{M} \pm Z_{\alpha/2}\sqrt{\frac{Y(M-Y)}{M^3}} \qquad \text{Eq B-2-8}$$

where

$\Re$ = design reliability or probability of success

Y = Number of Monte Carlo simulation points within the constraint boundary

M = Total number of Monte Carlo simulation points

$Z_{\alpha/2}$ = Standard Normal Quantile, e.g., 1.96 for α= 5%

α = Level of significance

This equation is most accurate for large M, and for $\Re$ near 0.5. As $\Re$ approaches 0 or 1, this approximation is less accurate, because the binomial distribution is not symmetrical (Conover, 1980). In the following discussion, Eq B-2-8 is termed method 1.

A slightly different equation is derived as follows (Hogg and Craig, 1978). By standard normalization, we have:

$$\frac{Y - M\Re}{\sqrt{M\Re(1-M\Re)}} \rightarrow N\{0,1\} \quad \text{as} \quad M \rightarrow \infty \qquad \text{Eq B-2-9}$$

And we can write a confidence interval for this as:

$$\Pr\left\{-Z_{\alpha/2} < \frac{Y - M\Re}{\sqrt{M\Re(1-M\Re)}} < Z_{\alpha/2}\right\} = 1-\alpha \qquad \text{Eq B-2-10}$$

or

$$\Pr\left\{\frac{(Y-M\Re)^2}{M\Re(1-M\Re)} < \left(Z_{\alpha/2}\right)^2\right\} = 1-\alpha \qquad \text{Eq B-2-11}$$

which can be solved for $\mathfrak{R}$ to give:

$$\mathfrak{R} = \frac{Z_{\alpha/2}^2 + 2Y \pm Z_{\alpha/2}\sqrt{Z_{\alpha/2}^2 + 4Y - 4\frac{Y^2}{M}}}{2\left(Z_{\alpha/2}^2 + M\right)} \qquad \text{Eq B-2-12}$$

The validity of Eq B-2-12 for $\mathfrak{R}$ near 0 and 1 is unknown. This equation is termed method 2.

Hogg and Craig (1978) also suggest a variable transformation method. The $\sin^{-1}$ of $\sqrt{Y/M}$ is essentially independent of $\mathfrak{R}$, with mean $\sin^{-1}\left(\sqrt{\mathfrak{R}}\right)$ and variance $M/4$. By approximate normal distribution, the confidence interval of this function is:

$$\Pr\left\{-Z_{\alpha/2} < \frac{\sin^{-1}\sqrt{Y/M} - \sin^{-1}\sqrt{\mathfrak{R}}}{\sqrt{M/4}} < Z_{\alpha/2}\right\} = 1-\alpha \qquad \text{Eq B-2-13}$$

Solving for $\mathfrak{R}$ gives:

$$\mathfrak{R} = \left[\sin\left(\sin^{-1}\sqrt{Y/M} + Z_{\alpha/2}\sqrt{M/4}\right)\right]^2 \qquad \text{Eq B-2-14}$$

This equation is termed method 3.

A two-level Monte Carlo simulation was run to test these confidence interval estimates. The inner loop ran a Conventional Monte Carlo simulation to estimate a probability and its confidence interval. This was compared with the true value to see if the confidence interval is correct. Then, the outer loop repeated this procedure many times to get the average failure rate. If the confidence interval is accurate, the average failre rate approaches α in the limit

200, 1000, and 25000 point Conventional Monte Carlo integrations were tested[1] with a uniform distribution in the range (0,1) with tabulations at <0.5, <0.6, <0.7, <0.8, <0.9, <0.95, <0.99, <0.995, <0.999. Both 95% and 99% confidence intervals were checked. The results are shown in Tables B-2-1,2,3, and Figure B-2-1.

[1] The 200 and 1000 point cases were repeated 100,000 times (method 0 only 30,000 times). The 25000 point case was repeated 20,300 times.

Table B-2-1
Monte Carlo Confidence Interval Test: 200 Point Simulation

Method	- - - - - - - Reliability % - - - - - - - - -								
	50	60	70	80	90	95	99	99.5	99.9
	95% Confidence Interval - Expected Value = 5 %								
Method 0	4.768	5.440	5.704	6.019	6.667	6.704	.1170	.044	.1220
Method 1	5.545	5.191	5.704	6.019	7.343	7.297	13.30	36.60	81.77
Method 2	5.545	5.191	5.460	4.301	4.389	3.184	5.267	8.004	1.804
Method 3	5.545	5.191	4.601	5.322	4.928	4.872	14.86	38.38	81.89
	99% Confidence Interval - Expected Value = 1 %								
Method 0	1.076	1.193	.9960	1.398	1.842	2.709	.003	.001	.0000
Method 1	1.272	1.193	1.141	1.398	1.842	2.648	13.19	36.56	81.77
Method 2	.8610	1.141	1.086	.9920	1.332	1.396	1.671	1.824	1.804
Method 3	.8610	.9220	1.141	1.057	1.113	1.171	13.30	36.60	81.77

Table B-2-2
Monte Carlo Confidence Interval Test: 1,000 Point Simulation

Method	- - - - - - - Reliability % - - - - - - - - -								
	50	60	70	80	90	95	99	99.5	99.9
	95% Confidence Interval - Expected Value = 5 %								
Method 0	5.053	4.950	4.907	4.980	5.517	5.553	7.280	12.72	0.057
Method 1	5.402	5.319	4.940	5.366	4.743	5.728	7.250	12.85	36.77
Method 2	5.402	4.931	4.893	5.353	5.130	4.974	3.676	3.812	8.185
Method 3	5.402	4.931	4.893	4.881	5.220	5.097	4.275	5.359	38.74
	99% Confidence Interval - Expected Value = 1 %								
Method 0	1.010	1.087	1.033	1.147	1.087	1.073	2.923	4.027	0.033
Method 1	1.030	.9750	1.035	.081	1.234	1.637	2.893	4.024	36.70
Method 2	1.030	1.072	.9400	1.037	1.015	1.080	.7720	1.369	2.038
Method 3	1.030	.9750	.9340	1.017	1.002	.8990	1.327	.8860	36.77

Table B-2-3
Monte Carlo Confidence Interval Test: 25,000 Point Simulation

Method	Reliability %								
	50	60	70	80	90	95	99	99.5	99.9
	95% Confidence Interval - Expected Value = 5 %								
Method 0	5.089	5.217	5.089	5.035	5.135	5.089	5.309	5.849	5.199
Method 1	5.019	5.162	5.370	5.079	5.354	5.153	5.005	4.882	5.108
Method 2	5.019	5.251	5.305	5.138	5.365	5.079	5.424	5.232	5.714
Method 3	5.084	5.148	5.207	5.064	5.330	4.995	5.517	5.468	5.734
	99% Confidence Interval - Expected Value = 1 %								
Method 0	1.062	0.851	1.153	1.062	1.172	1.016	1.099	1.291	1.373
Method 1	1.025	0.852	1.044	0.951	1.020	0.985	0.921	1.060	2.286
Method 2	1.025	0.867	1.059	0.941	1.025	1.005	1.040	0.970	1.252
Method 3	1.010	1.030	1.020	1.089	1.034	1.010	.9655	.9360	1.030

All estimation methods improve as the number of Monte Carlo points increase. For higher reliabilities, more Monte Carlo points are required for accurate results. Methods 0 and 1 are clearly inferior. Methods 2 and 3 are fairly equivalent, however, method 2 is most accurate at high reliabilities (>95%), requiring fewer Monte Carlo points.

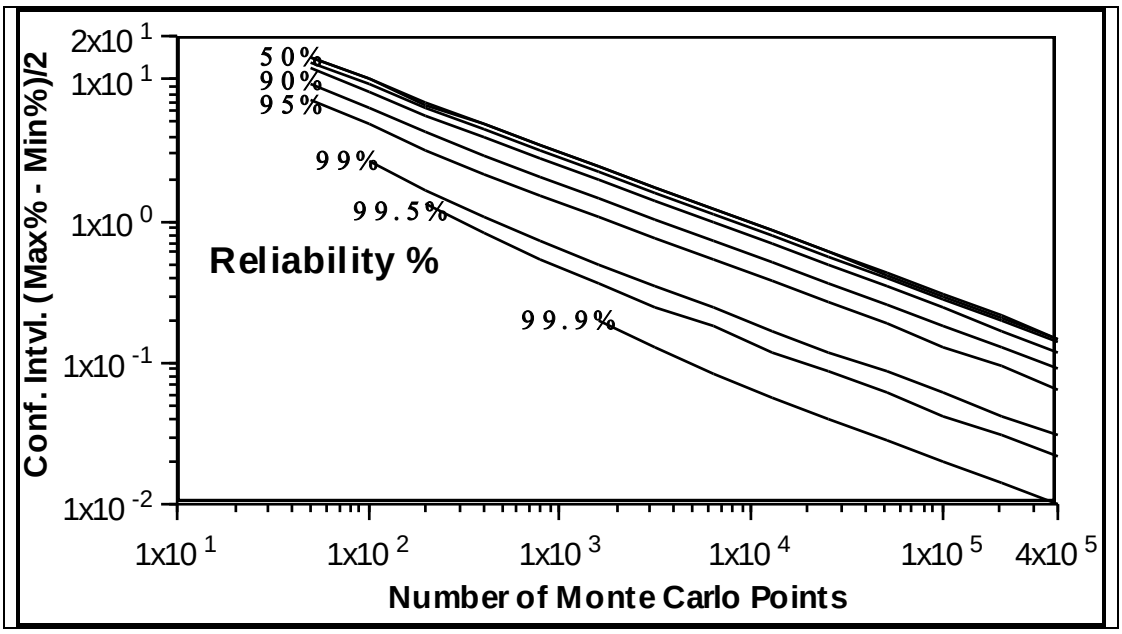
Method 2 is preferred, and will be used henceforth.

If accurate confidence intervals are necessary, 200 Monte Carlo points are sufficient for reliabilities up to 90%. 1,000 Monte Carlo points are required when reliabilities are 90%-99%. For reliabilities greater than 99%, more than 1,000 Monte Carlo integration points are required. However, it should be recognized that a confidence interval at 99% reliability is much smaller than the confidence interval at 90% reliability, for the same number of Monte Carlo points. So even though the confidence interval might be inaccurate at 99% reliability, the error is relatively small.

For convenience, Table B-2-5 and Figure B-2-2 provide the 95% confidence intervals for different reliability/point combinations. For 50 and 100 Monte Carlo integration points, the exact binomial confidence interval is given. For 200 points and greater, method 2 (Eq B-2-12) is used. Note that values are given as (Max%-Min%)/2, since the confidence interval is *not* symmetrical.

Table B-2-5 and Figure B-2-2
95% Confidence Intervals for Conventional Monte Carlo Integration as a Function of Reliability and Number of Simulation Points

# Points	- - - - - - - Reliability % - - - - - - - -								
	50	60	70	80	90	95	99	99.5	99.9
	Values are (Max%-Min%)/2. *Not* Symmetrical								
50	14.47	14.21	13.38	11.84	9.24	7.13			
100	10.17	9.98	9.37	8.26	6.36	4.82	2.71		
200	6.86	6.73	6.30	5.52	4.19	3.11	1.65	1.34	
400	4.88	4.78	4.47	3.91	2.95	2.17	1.08	.83	
800	3.46	3.39	3.17	2.77	2.08	1.52	.73	.54	
1,600	2.45	2.40	2.24	1.96	1.47	1.07	.50	.36	.20
3,200	1.73	1.70	1.59	1.39	1.04	.76	.35	.25	.13
6,400	1.23	1.20	1.12	.98	.74	.53	.25	.18	.083
12,800	.87	.85	.79	.69	.52	.38	.17	.12	.057
25,600	.61	.60	.56	.49	.37	.27	.12	.086	.040
51,200	.43	.42	.40	.35	.26	.19	.087	.061	.028
102,400	.31	.30	.28	.25	.18	.13	.061	.043	.020
204,800	.22	.21	.20	.17	.13	.094	.043	.031	.014
409,600	.15	.15	.14	.12	.091	.066	.031	.022	.010



A more stringent statistical approach is to use the Chebyshev inequality (Hogg and Craig, 1978), which for the binomial distribution is:

$$Pr\left[\left| \frac{Y}{M} - \Re \right| \geq \in \right] \leq \frac{\Re(1-\Re)}{M \in^2} \qquad \text{Eq B-2-15}$$

where $\in$ is the specified accuracy requirement in $\leftarrow$

If one must know the true reliability to within ±1% accuracy with 95% confidence, this is:

$$\Pr\left[\left| \frac{Y}{M} - \Re \right| \geq 0.01 \right] \leq 0.05 \qquad \text{Eq B-2-16}$$

If the true reliability is 90%, this requires at least ($0.9 \infty 0.1 / 0.05 / 0.01^2$) = 18,000 Monte Carlo points, i.e.:

$$M^{Min} \leq \frac{\Re(1-\Re)}{\alpha \varepsilon^2} \qquad \text{Eq B-2-17}$$

The only difficulty is that $\leftarrow$ must be known beforehand, which is impossible. However, this can be estimated from Y/M, as the Monte Carlo simulation progresses.

To be conservative, we should use $\leftarrow$=Y/M$-\in$ in Eq B-2-17, since reducing $\leftarrow$ maximizes $\leftarrow(1-\leftarrow)$ (assuming $\leftarrow$>50%). However, even this is incorrect α% of the time (α/2% if the Monte Carlo simulation distribution is symmetrical). So technically, Eq B-2-17 using $\leftarrow$=Y/M$-\in$ is incorrect up to α% of the time. Thus, the confidence interval could be as low as $(1-\alpha)^2$. I suspect that this is overly conservative however.

I ran a Monte Carlo simulation to test this. 100, 200, 500, 1000, 2500, 5000, 10000, 25000, 50000, and 100000 Monte Carlo sets were generated with a (0,1) uniform probability distribution, and tabulated in the probability ranges <0.5, <0.6, <0.7, <0.8, <0.9, <0.95, <0.99, <0.995, <0.999. 95% and 99% confidence intervals were calculated for the Chebyshev inequality, using (1) zero correction, (2) 1/2 $\in$ correction, and (2) 1 $\in$ correction to the estimated probability. This was repeated 8,150 times.

Failure rates (incorrect confidence intervals) were extremely low. The 99% confidence interval was never violated, and the 95% confidence interval was only violated a handful of times. The highest failure rate was only 0.3%. Thus, we may conclude that the Chebyshev inequality is very conservative, and therefore not useful for this work, since tighter confidence intervals can be obtained from the method 2 approximation, as shown in Table B-2-6.

Table B-2-6

Required Number of Monte Carlo Integration Points for Specified 95% Confidence Intervals: Comparison of Chebyshev Inequality and Best Estimation Method (Eq B-2-12)

Probability %	Required Accuracy % (± 95% Conf. Intvl.)	Min. Req'd M Chebyshev	M Eq B-2-12
90	± 1.0	19,600	3,460
90	0.1	1,960,000	346,000
95	1.0	11,300	1,820
95	0.1	968,000	182,000
99	0.1	217,600	38,000
99	0.03	2,270,000	423,000
99	0.01	20,000,000	3,800,000
99.9	0.03	289,000	42,600
99.9	± 0.01	2,200,000	384,000

B.3. Statistical Likelihood that a Conventional Monte Carlo Run is Within ±X% Accuracy, Given M and ←

This section develops a statistical calculation required for the Conventional Monte Carlo and boundary-approximation comparison (section V.C.2.c): Given that the boundary-approximation procedure is within ±x% accuracy, what are the odds that a Conventional Monte Carlo run will be more accurate?

Section V.C.2.c. requires the following statistical calculation: What is the statistical likelihood that a Conventional Monte Carlo run is within ±x% of the true reliability, given the number of Monte Carlo points? The true reliability is assumed known, however it must be estimated in practice.

The problem is stated: What is the probability that a single Conventional Monte Carlo run is within ±x%, given ← and M? This may be calculated in two steps: (1) Find the values of Y that are within this accuracy. (2) Calculate the probability that these values of Y occur. That is:

$$\mathrm{Y}^{\mathrm{Min}} = int\left\{\left(\Re - \mathrm{x}\%\right)\cdot \mathrm{M}\right\} + 1 \qquad \text{Eq B-3-3a}$$

$$\mathrm{Y}^{\mathrm{Max}} = int\left\{\left(\Re + \mathrm{x}\%\right)\cdot \mathrm{M}\right\} \qquad \text{Eq B-3-3b}$$

where

$\mathrm{Y}^{\mathrm{Min}}$ is the minimum value of Y that satisfies the accuracy (within -x% of ←)

$\mathrm{Y}^{\mathrm{Max}}$ is the maximum value of Y that satisfies the accuracy (within +x% of ←)

$int\{\cdot\}$ is the greatest integer not greater than the argument

Now the probability that a Conventional Monte Carlo run is within ±x% accuracy is the probability that Y is between Y^{Min} and Y^{Max}, which is:

$$\Pr\{\text{Conv.MC Error} < x\%\} = \Pr\{Y^{Min} \le Y \le Y^{Max}\}$$
$$= \sum_{Y=Y^{Min}}^{Y^{Max}} \Pr\{b(M,\Re) = Y\} = \sum_{Y=Y^{Min}}^{Y^{Max}} \binom{M}{Y} \Re^{Y} (1-\Re)^{M-Y} \qquad \text{Eq B-3-4}$$

Eq B-3-4 provides the probability that a Conventional Monte Carlo run is within ±x% accuracy. However, section V.C.2.c. requires the statistical likelihood or odds that the boundary-approximation procedure is more accurate. This is calculated as:

$$\Pr\{\text{B.A. is more accurate}\} = \frac{\Pr\{\text{Conv.MC Error} < x\%\}}{\Pr\{\text{Conv.MC Error} > x\%\}} = \frac{\Pr\{\text{Conv.MC Error} < x\%\}}{1 - \Pr\{\text{Conv.MC Error} < x\%\}} \qquad \text{Eq B-3-5}$$

If M is low and x% is small, it might happen that *no* values of Y satisfy the accuracy requirement, i.e., $Y^{Min} > Y^{Max}$. This means that Conventional Monte Carlo cannot be as accurate as ±x%.

Appendix C. Distillation Test Problems

This section provides documentation for the distillation test problems. Problem specifications, database information, column profiles, and constraint boundary plots are given.

Three binary distillation test problems were developed: (1) Acetone-Benzene, (2) Acetone-Water, (3) 2-Methyl-1-Butene – Isoprene (2MB1-IPM). One ternary distillation test problem was developed: Acetone-Benzene-Toluene.

For all test problems, feed flowrate, tray efficiency, and thermodynamic database were uncertain. Values for feed flowrates and tray efficiency uncertainties were set arbitrarily. Thermodynamic database uncertainties were calculated by regressing several data sets to get reasonable 'ballpark' values for these uncertainties.

The thermodynamic database model used Modified Wilson (Tsuboka and Katayama, 1975) for the liquid phase and the ideal gas law for the vapor phase. All K-value uncertainties were accounted for in the Modified Wilson parameters.

Component vapor pressure was modeled by the Miller equation (Miller, 1964). Coefficient values are listed in Table C-1.

Table C-1
Coefficient Values for Miller Equation Vapor Pressure[1]

	Acetone	Benzene	Water	Toluene	2MB1	IPM
A	-5.0010E+03	-5.3903E+03	-6.4437E+03	-5.9164E+03	-4.3628E+03	-4.0872E+03
B	2.5978E+01	2.6678E+01	2.7939E+01	2.7109E+01	2.4640E+01	2.1671E+01
C	-2.5492E-02	-2.6122E-02	-2.1853E-02	-2.505E-02	-2.5079E-02	-1.5552E-02
D	2.0511E-05	1.9435E-05	1.5097E-05	1.7200E-05	2.1040E-05	1.0914E-05

Vapor and liquid enthalpies were modeled by a 3'rd order polynomial curve-fit with ideal mixing. The curves were either fitted to pseudo-data, generated by the HYSIM process simulator, or from available physical property data. Table C-2 provides values of the database constants that were used:

[1] Equation is: $ln\left(P_i^{sat}\right) = \frac{A_i}{T} + B_i + C_i T + D_i T^2$; where T is the temperature in degK, and P_i^{sat} is the component vapor pressure in kPA.

Table C-2
Coefficient Values for Enthalpy Model[1]

	Acetone	Benzene	Water	Toluene	2MB1	IPM
Liquid						
A	-1.7352E+02	-1.3282E+02	-3.1572E+01	-1.3687E+01	8.9103E+02	8.8508E+02
B	4.7899E-01	3.4724E-01	9.9455E-01	4.2771E-01	4.8025E-01	4.7203E-01
C	1.5607E-04	2.4836E-04	9.7403E-07	0.0000E+00	3.9901E-04	3.2067E-04
D	2.6212E-07	3.6616E-08	7.5758E-08	0.0000E+00	1.9134E-07	3.1226E-07
Vapor						
A	6.9914E+01	6.0663E+01	1.0653E+03	1.7580E+02	1.0634E+03	1.0649E+03
B	2.6776E-01	2.0793E-01	3.6318E-01	2.2650E-01	3.1344E-01	3.0447E-01
C	2.2912E-04	2.8843E-04	5.4870E-04	2.7949E-04	3.2717E-04	3.0959E-04
D	-4.4697E-08	-7.3737E-08	-1.7677E-06	-6.9459E-08	-7.3310E-07	-6.7330E-07

The liquid solution model requires liquid molar volumes for each component. The Hankinson-Thomson correlation (Hankinson and Thomson, 1979) was used, see section IV.B.2.c. Table C-3 lists the characteristic volumes for each component.

Table C-3
Characteristic Volumes for Hankinson-Thomson Correlation

	Acetone	Benzene	Water	Toluene	2MB1	IPM
$\frac{cm^3}{gmol}$	208.4	254.57	54.232	313.708	288.7	267.8

C-1. Acetone-Benzene Distillation

Acetone-benzene is a relatively ideal, medium-wide boiling system. Relative volatilities range from 1.6 to 2.8. Simulation convergence is fast without any difficulties.

The test problem was specified as a separation of the feed mixture into two low purity component product streams.

C-1.a. Problem Description

[1] Equation is: $H_i = A_i + B_i T + C_i T^2 + D_i T^3$; where T is the temperature in degF, and H_i is the component enthalpy in Btu/lbm. The stream enthalpy is the mass average.

The distillation column has 25 stages (including condenser), with the feed to stage 15, and a total condenser operating at atmospheric pressure. Vapor overhead pressure drop is 3 psi. Pressure drop per stage is 0.1 psi/stage.

Feed is nominally 500/500 lbmol/hr (acetone/benzene) and a saturated liquid at 20 psia.

Two product specifications were set initially: (1) 80% minimum acetone overhead recovery, and (2) 95 mol% minimum acetone overhead purity. Simulation gave a reboiler vapor of 100,028 lbm/hr, with 17.3 mol% acetone in the bottoms.

A vapor rate safety factor was then added so the column floods at 125,000 lbm/hr, and a maximum acetone bottoms impurity was specified as 25 mol%. Of the four inequality constraint specifications, flood point and acetone recovery were used for control variables. This gives:

Table C-1.a-1
Acetone-Benzene Test Problem Inequality Constraints

CV-1)	Reboiler Vapor Rate ≤ 125,000 lbm/hr
CV-2)	Acetone Overhead Recovery ≥ 80%
A)	Acetone Overhead Purity ≥ 95 mol%
B)	Acetone Bottoms Impurity ≤ 25 mol%

The following statistical uncertainties were used for feed flowrate and tray efficiency:

Table C-1.a-2
Acetone-Benzene Test Problem Feed Flowrate and Tray Efficiency Uncertainties

- Acetone feed flowrate is normally distributed with a mean of 500 lbmol/hr and a standard deviation of 100 lbmol/hr.
- Benzene feed flow is normally distributed with a mean of 500 lbmol/hr and a standard deviation of 100 lbmol/hr.
- Tray efficiency is uniformly distributed with a minimum of 50% and a maximum of 60%.

C-1.b. Converged Run listing

```
Number of Stages =  25
Total  Condenser   - Partial Reboiler
Feed stream feed            to stage  15
Draw stream ovd             from stage  1
```

Draw stream btms from stage 25
Murphree vapor efficiency = 60.00 %
Condenser Pressure = 15.00 psia
Vapor Ovd Pressure = 18.00 psia
Reboiler Pressure = 20.30 psia
Top stage: Recovery = .8000 for component 1
Top stage 2: Component mole fraction = 0.95 for component 1
Convergence criterion = 5.0000E-07
Convergence option = fractional percentage

Feed Streams for Operation

feed to stage 15
Temp (deg F): 162.43672 MolWt: 68.09750
Press (psia): 20.00000 Spec: P,Tbub
Quality: .00 H (Btu/lbmol): -5350.0273

Component	LbMoleFlow	Mole %	LbMassFlow	Mass %
Acetone	500.0000	50.0000	29040.00	42.6447
Benzene	500.0000	50.0000	39057.50	57.3553
Total	1000.000	100.0000	68097.50	100.0000

Draw Streams for Operation

ovd from stage 1
Temp (deg F): 135.00193 MolWt: 59.08175
Press (psia): 15.00000 Spec: None
Quality: .00 H (Btu/lbmol): -6131.3340

Component	LbMoleFlow	Mole %	LbMassFlow	Mass %
Acetone	400.0000	95.0000	23232.00	93.3893
Benzene	21.05246	5.0000	1644.513	6.6107
Total	421.0525	100.0000	24876.51	100.0000

btms from stage 25
Temp (deg F): 178.51074 MolWt: 74.65441
Press (psia): 20.29999 Spec: None
Quality: .00 H (Btu/lbmol): -4870.9633

Component	LbMoleFlow	Mole %	LbMassFlow	Mass %
Acetone	100.0000	17.2727	5808.000	13.4379
Benzene	478.9475	82.7273	37412.98	86.5621
Total	578.9475	100.0000	43220.98	100.0000

Stages for Operation

Stg	Press	Temp	Vap-lbm	Liq-lbm	Feed	LiqDraw	VapDraw	Heat-Btu/hr
1	15.00	135.00	0.0	61335.4	.0	24876.5	.0	-.1840E+06
2	18.00	145.20	86211.9	63069.8	.0	.0	.0	0.0000E+00
3	18.10	145.89	87946.3	63352.0	.0	.0	.0	0.0000E+00
4	18.20	146.61	88228.5	63654.1	.0	.0	.0	0.0000E+00
5	18.30	147.37	88530.6	63977.5	.0	.0	.0	0.0000E+00
6	18.40	148.17	88854.0	64323.4	.0	.0	.0	0.0000E+00
7	18.50	149.01	89199.9	64693.0	.0	.0	.0	0.0000E+00
8	18.60	149.91	89569.5	65087.3	.0	.0	.0	0.0000E+00
9	18.70	150.87	89963.8	65507.0	.0	.0	.0	0.0000E+00
10	18.80	151.89	90383.5	65952.3	.0	.0	.0	0.0000E+00
11	18.90	152.98	90828.9	66423.0	.0	.0	.0	0.0000E+00
12	19.00	154.14	91299.5	66917.7	.0	.0	.0	0.0000E+00
13	19.10	155.38	91794.2	67433.8	.0	.0	.0	0.0000E+00
14	19.20	156.70	92310.3	67871.8	.0	.0	.0	0.0000E+00
15	19.30	158.72	92748.3	135724.3	68097.5	.0	.0	0.0000E+00
16	19.40	159.17	92503.3	135936.0	.0	.0	.0	0.0000E+00
17	19.50	159.68	92715.0	136214.7	.0	.0	.0	0.0000E+00
18	19.60	160.30	92993.7	136588.3	.0	.0	.0	0.0000E+00
19	19.70	161.05	93367.3	137095.1	.0	.0	.0	0.0000E+00
20	19.80	162.00	93874.1	137786.7	.0	.0	.0	0.0000E+00
21	19.90	163.26	94565.7	138730.2	.0	.0	.0	0.0000E+00
22	20.00	164.94	95509.2	140006.8	.0	.0	.0	0.0000E+00
23	20.10	167.25	96785.8	141701.9	.0	.0	.0	0.0000E+00
24	20.20	170.45	98480.9	143249.1	.0	.0	.0	0.0000E+00
25	20.30	178.51	100028.1	0.0	.0	43221.0	.0	0.1846E+06

Stage Composition and Relative Volatilities for Operation

Stg	Acetone Liquid Mole %	K-value Acetone	K-value Benzene	Relative Volatility
2	93.35%	1.026	.6327	1.622
3	91.56%	1.034	.6323	1.635
4	89.64%	1.043	.6319	1.650
5	87.57%	1.052	.6315	1.666
6	85.35%	1.063	.6312	1.685
7	82.95%	1.076	.6311	1.705
8	80.37%	1.090	.6310	1.728
9	77.61%	1.106	.6312	1.753
10	74.66%	1.125	.6316	1.781
11	71.52%	1.146	.6323	1.813
12	68.19%	1.171	.6335	1.848
13	64.69%	1.199	.6352	1.888
14	61.04%	1.231	.6376	1.931
15	55.21%	1.290	.6426	2.007
16	54.72%	1.295	.6433	2.013
17	54.01%	1.303	.6442	2.023
18	53.00%	1.314	.6454	2.037
19	51.57%	1.331	.6473	2.057
20	49.54%	1.356	.6501	2.086
21	46.72%	1.394	.6545	2.130
22	42.84%	1.452	.6617	2.194
23	37.64%	1.541	.6736	2.287

24	30.98%	1.681	.6942	2.422
25	17.27%	2.119	.7664	2.765

C-1.c. Database Regression

Gmehling and Onken (1977ff) have several sets of acetone-benzene vapor-liquid equilibria data, which include both TPx and TPxy measurements. For simplicity, I took the two isobaric TPxy sets at 760 mmHG which passed the consistency tests of Gmehling and Onken (1977ff), and regressed them on the SYNDES simulator (SYNDES, 1987) assuming they were TPx only. Figure C-1.c-1 plots the calculated-measured residuals in pressure.

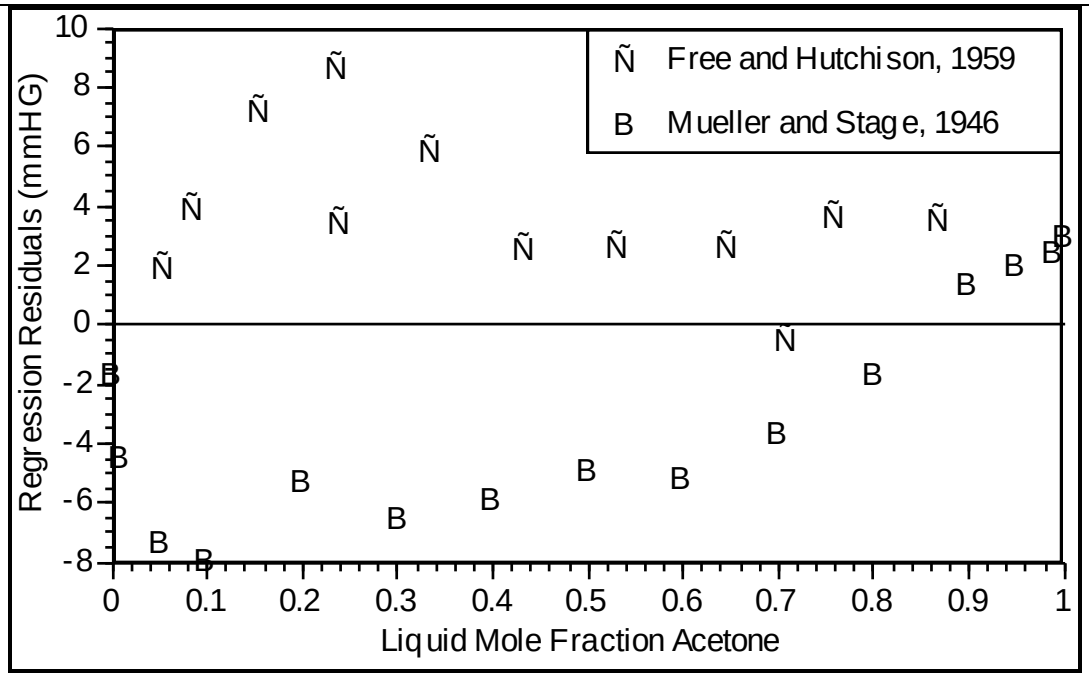


Figure C-1.c-1
Residual Pressure Plot of Modified Wilson Regression of Acetone-Benzene VLE Data

Obviously, there is some discrepancy between the two data sets, and further analysis is warranted. However, for our purposes, any reasonable values will do.

The Modified Wilson coefficients and their uncertainties used for this work are shown in Table C-1.c-1:

Table C-1.c-1
Acetone - Benzene Test Problem: Statistical Distribution for Modified Wilson Solution Model Coefficients

1-Acetone 2-Benzene	Bivariate Normal Distribution	
	Mean cal/gmole	Standard Deviation
$\lambda_{12} - \lambda_{11}$	577.471	49.706
$\lambda_{21} - \lambda_{22}$	-234.411	31.409
	Correlation Coefficient	-0.99279935

Or writing the covariance matrix:

$$Covar\begin{pmatrix}\lambda_{12}-\lambda_{11}\\ \lambda_{21}-\lambda_{22}\end{pmatrix} = \begin{pmatrix} 2470.686 & -1549.974 \\ -1549.974 & 986.525 \end{pmatrix}$$

C-1.d. Constraint Boundary Mapping

This section presents pictures of the constraint boundary in the various dimensions. First the feed-space, then the database-space is plotted.

C-1.d.i. Feed Flowrate Parameter Space

Figure C-1.d.i-1 plots the constraint boundary in the feed flowrate dimensions (p=2). The database is set at its design (or mean) value. The tray efficiency is set at both maximum and minimum values, with both boundary lines shown. Standard deviation contours are drawn for the feed flowrate uncertainties.

The top boundary line is the overhead purity. The bottom boundary line is the bottoms impurity. Overhead purity is sensitive to tray efficiency, while bottoms impurity is not. Note that while the boundary is open-ended towards the left, a practical limit exists at zero flowrate.

Note that the boundary is both concave and convex. Also, it has long linear regions with one sharp corner where the two constraints meet.

Next, the database was set at +3 and -3 standard deviations along the principal axis (top-left, and bottom-right) and the feed-space boundary mapped at 50% and 60% tray efficiency. This plot is shown in Figure C-1.d.i-2. Note that while the database appears to have a significant impact on the constraint boundary, the impact is offsetting — the top-left quadrant of the database values shrinks the constraint boundary while the bottom-right quadrant enlarges the boundary. Like the tray efficiency uncertainty, the overhead purity constraint is sensitive to the database uncertainties, while the bottoms impurity is not.

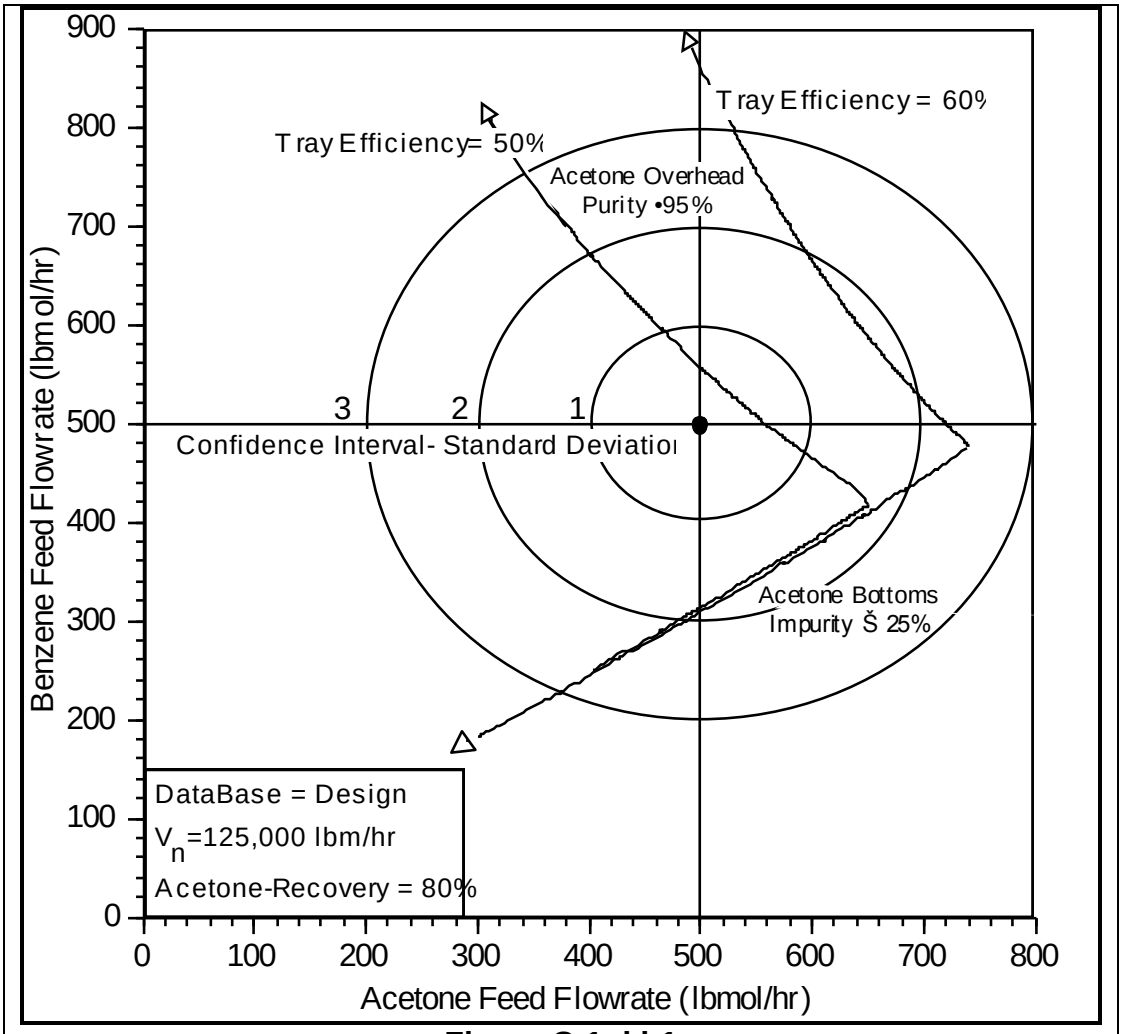


Figure C-1.d.i-1
Constraint Boundary in the Component Feed Flowrate Space, Showing Minimum and Maximum Tray Efficiency Sensitivity

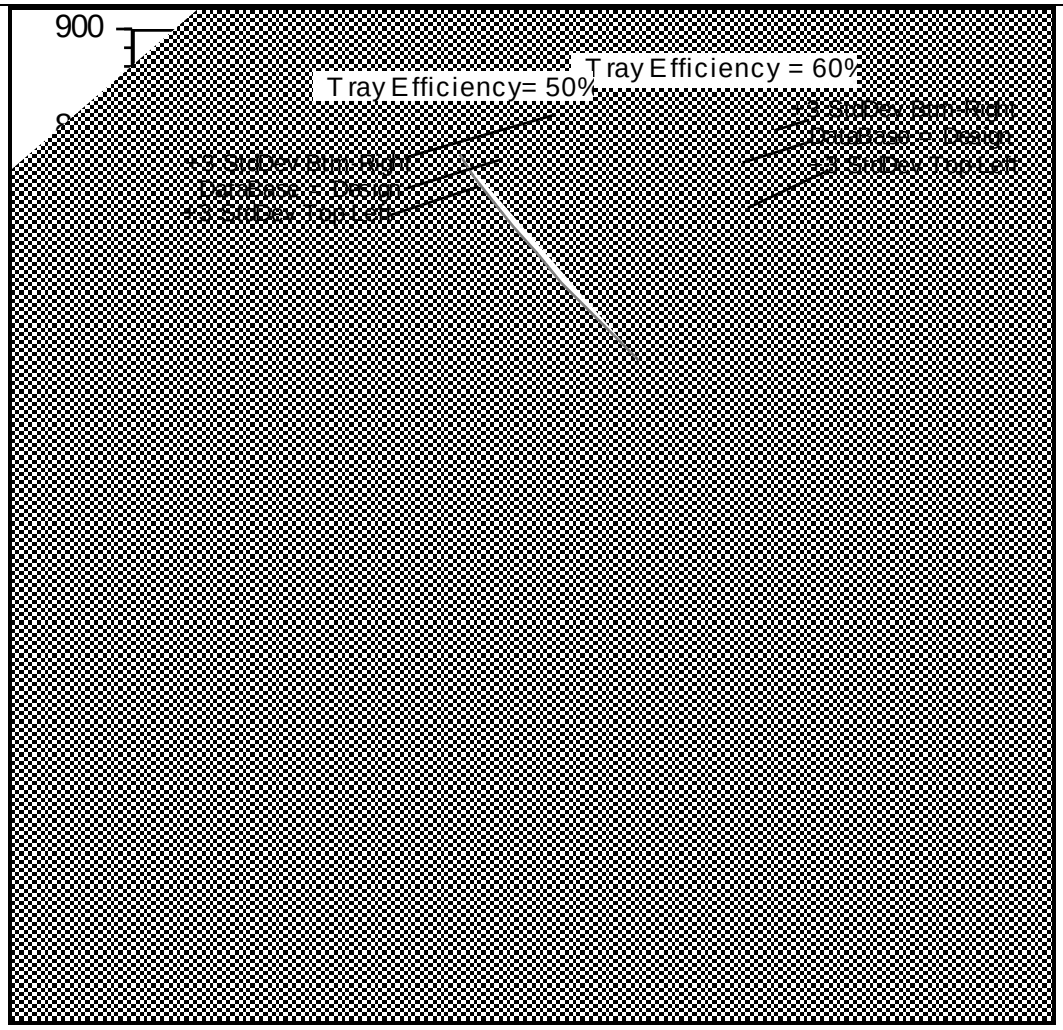


Figure C-1.d.i-2
Constraint Boundary in the Component Feed Flowrate Space, Showing Thermodynamic DataBase Sensitivity

C-1.d.ii. Liquid Activity Coefficients Parameter Space

For a binary distillation, the database has two uncertain parameters (λ_{12} and λ_{21}). The constraint boundary in λ-space is shown in Figure C-1.d.ii-1. All other parameters are at their design value.

Note that the constraint boundary is 'ridge-shaped' and open-ended in two directions.[1] However, the 95% confidence interval for the database is tiny in comparison.

[1] It is not known for certain whether or not this boundary is open-ended, because the search was stopped due to convergence problems. This region is vanishingly improbable, however.

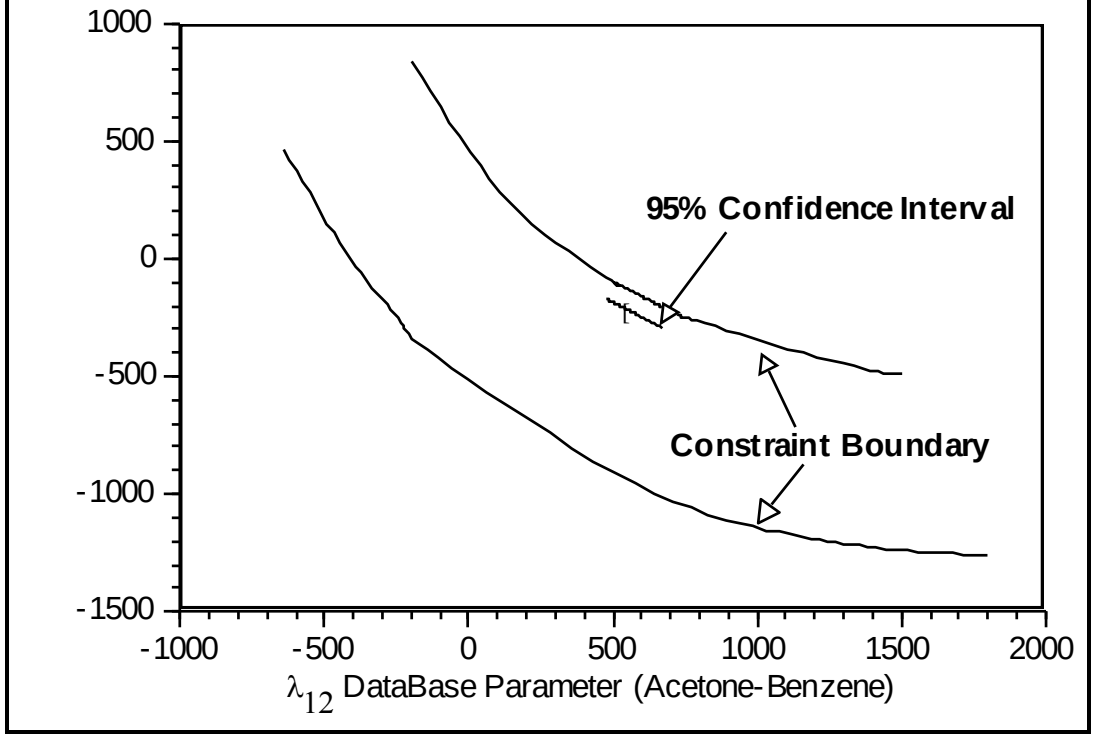


Figure C-1.d.ii-1
Map of Constraint Boundary in λ-Space

Figure C-1.d.ii-2 plots the constraint inequality residual as a function of the database parameters along the uncertainty principal axis from -5 to +5 standard deviations. This is shown for: (1) The design point[1]; (2) The design point with minimum tray efficiency (50%); (3) The acetone feed flowrate maximum axial boundary point; (4) The benzene feed flowrate maximum axial boundary point; (3) The benzene feed flowrate minimum axial boundary point. The plot shows that the inequality constraint has a quadratic shape with respect to the database parameters, but it is neither constant nor symmetrical over the region.

[1] The true value of the constraint residual at the design point is negative. It is shifted to zero so that the slope comparison may be highlighted.

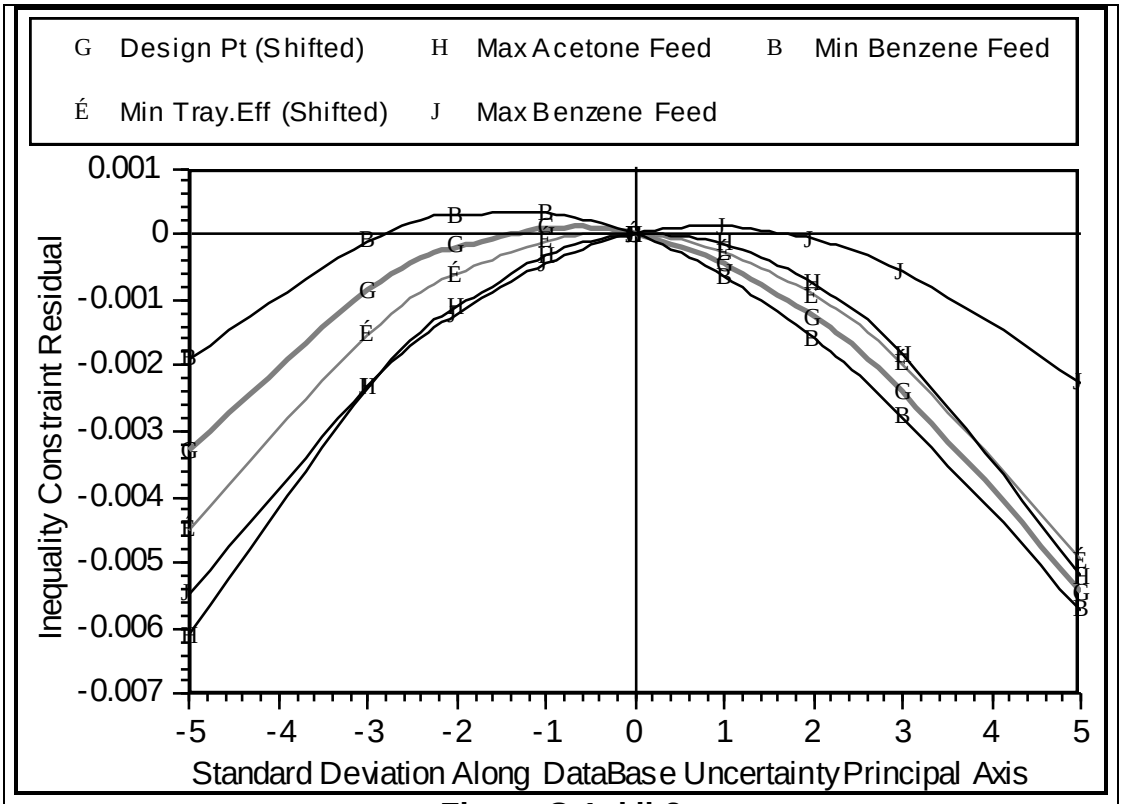


Figure C-1.d.ii-2
Shape of the Inequality Constraint as a Function of the Database Uncertainty Parameters at the Design Point and Other Constraint Boundary Points

C-2. Acetone-Water Distillation

Acetone-water was chosen as a wide boiling, highly nonideal system. Relative volatilities range from 2 to 47. Nonideality is so great that convergence is very difficult and the iteration had to be heavily damped (damper of 0.2-0.3), which greatly increased execution time. The test problem was specified as a high purity stripping of acetone from water.

C-2.a. Problem Description

The distillation column has 14 stages (including condenser), with the feed to stage 8, and a total condenser operating at atmospheric pressure. Vapor overhead pressure drop is 3 psi. Pressure drop per stage is 0.1 psi/stg.

Acetone/water feed was nominally 50/950 lbmol/hr, and a saturated liquid at 22 psia. The column uses a high purity specification of at least 99.95 mol% water for the bottoms product stream. Tray efficiency is low (40%), as we would expect for a wide-boiling system. Maximum allowable acetone in the bottoms is 25 lbm/hr. Maximum allowable water in the overhead is 10 lbmol/hr. The latter specification was set as one of the control variables. The

other control variable was setting the reboiler heat input to $3 * 10^6$ Btu/hr, which is a 12.5% safety factor in the vapor boilup rate.

Table C-2.a-1
Acetone-Water Test Problem Inequality Constraints

CV-1)	Reboiler Heat Input ≤ 3e6 Btu/hr
CV-2)	Water Overhead ≤ 10 lbmol/hr
A)	Acetone Overhead Purity ≥ 99.95 mol%
B)	Acetone Bottoms Flowrate ≤ 25 lbm/hr

The following statistical uncertainties were used for the feed flowrate and tray efficiency:

Table C-2.a-2
Acetone-Water Test Problem Feed Flowrate and Tray Efficiency Uncertainties

- Acetone feed flow is normally distributed with a mean of 50 lbmol/hr and a standard deviation of 10 lbmol/hr.
- Water feed flow is normally distributed with a mean of 950 lbmol/hr and a standard deviation of 10 lbmol/hr.
- Tray efficiency is uniformly distributed with a minimum of 30% and a maximum of 40%.

C-2.b. Converged Run listing

```
---------------------------------------------------------------------------
Number of Stages =  14
Total  Condenser   - Partial Reboiler
Feed stream feed              to stage  8
Draw stream ovd               from stage  1
Draw stream btms               from stage  14
Murphree vapor efficiency =  40.00 %
Condenser Pressure =  15.00 psia
Vapor Ovd Pressure =  18.00 psia
Reboiler  Pressure =  19.20 psia
Top stage:  Component Molar Flow Rate in Liquid  di =  10.00 for component  2
Bottom stage:  Component Mole Fraction = 0.9995 for component 2
Convergence criterion       =  1.0000E-06
Convergence option   =   fractional percentage
```

Feed Streams for Operation

feed to stage 8
Temp (deg F): 185.86380 MolWt: 20.01825
Press (psia): 22.00000 Spec: P,Tbub
Quality: .00 H (Btu/lbmol): 2407.3150

Component	LbMoleFlow	Mole %	LbMassFlow	Mass %
Acetone	50.00000	5.0000	2904.000	14.5068
Water	950.0000	95.0000	17114.25	85.4932
Total	1000.000	100.0000	20018.25	100.0000

Draw Streams for Operation

ovd from stage 1
Temp (deg F): 135.67443 MolWt: 51.34980
Press (psia): 15.00000 Spec: None
Quality: .00 H (Btu/lbmol): -4760.8442

Component	LbMoleFlow	Mole %	LbMassFlow	Mass %
Acetone	49.53018	83.2018	2876.713	94.1067
Water	10.00000	16.7982	180.1500	5.8933
Total	59.53018	100.0000	3056.863	100.0000

btms from stage 14
Temp (deg F): 224.58500 MolWt: 18.03502
Press (psia): 19.19999 Spec: None
Quality: .00 H (Btu/lbmol): 3468.0750

Component	LbMoleFlow	Mole %	LbMassFlow	Mass %
Acetone	.4698220	.0500	27.28726	.1609
Water	940.0001	99.9500	16934.10	99.8391
Total	940.4699	100.0000	16961.39	100.0000

Stages for Operation

Stg	Press	Temp	Vap-lbm	Liq-lbm	Feed	LiqDraw	VapDraw	Heat-Btu/hr
1	15.00	135.67	0.0	3078.6	.0	3056.9	.0	0.3262E+06
2	18.00	146.20	6135.5	2887.5	.0	.0	.0	0.0000E+00
3	18.10	147.85	5944.4	2579.2	.0	.0	.0	0.0000E+00
4	18.20	150.14	5636.0	2193.4	.0	.0	.0	0.0000E+00
5	18.30	153.60	5250.3	1703.9	.0	.0	.0	0.0000E+00
6	18.40	160.52	4760.7	1120.4	.0	.0	.0	0.0000E+00
7	18.50	179.12	4177.3	879.4	.0	.0	.0	0.0000E+00
8	18.60	181.34	3936.2	20317.6	20018.3	.0	.0	0.0000E+00
9	18.70	185.58	3356.2	20037.8	.0	.0	.0	0.0000E+00
10	18.80	191.90	3076.4	19735.6	.0	.0	.0	0.0000E+00
11	18.90	200.69	2774.2	19473.5	.0	.0	.0	0.0000E+00
12	19.00	210.66	2512.1	19317.8	.0	.0	.0	0.0000E+00
13	19.10	218.67	2356.4	19263.2	.0	.0	.0	0.0000E+00
14	19.20	224.58	2301.8	0.0	.0	16961.4	.0	0.2665E+07

Stage Composition and Relative Volatilities for Operation

Stg	Acetone Liquid Mole %	K-value Acetone	K-value Water	Relative Volatility
1	83.20%	1.092	.5435	2.009
2	76.49%	1.154	.5006	2.304
3	67.76%	1.272	.4292	2.963
4	55.90%	1.500	.3665	4.092
5	39.21%	2.064	.3136	6.583
6	16.53%	4.600	.2872	16.01
7	4.42%	13.909	.4026	34.55
8	3.99%	14.974	.4198	35.67
9	3.24%	17.242	.4570	37.73
10	2.36%	20.928	.5194	40.29
11	1.46%	26.661	.6197	43.02
12	0.74%	34.077	.7537	45.21
13	0.31%	40.759	.8772	46.46
14	0.05%	46.097	.9775	47.16

C-2.c. Database Regression

For the regression, I used six 760 mmHG TPxy (regressed as TPx) data sets that passed the Gmehling and Onken (1977ff) consistency tests. Figure C-2.c-1 plots the calculated-measured pressure residuals.

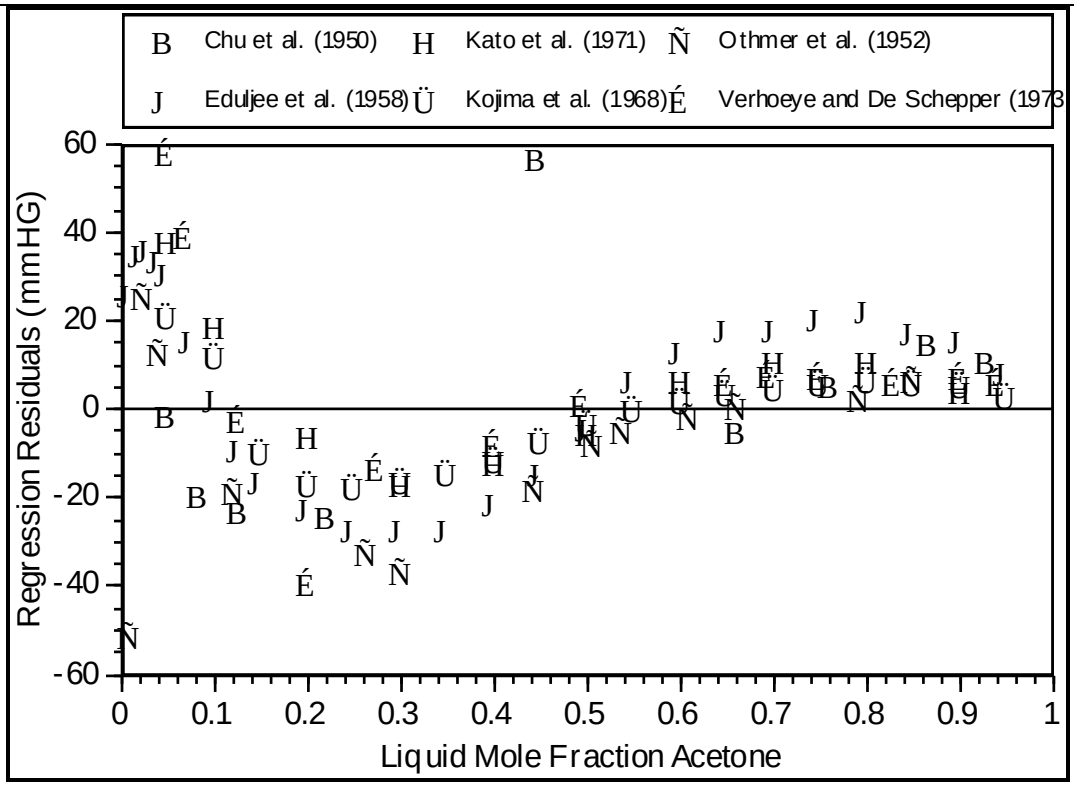


Figure C-2.c-1
Residual Plot of Modified Wilson Regression of Acetone-Water VLE Data

Obviously, there is systematic error in the fit and some very poor data points. This was ignored. The statistical distribution of the solution model coefficients used for this work are shown in Table C-2.c-1.

Table C-2.c-1
Acetone - Water Test Problem: Statistical Distribution for Modified Wilson Solution Model Coefficients

	Bivariate Normal Distribution	
1-Acetone 2-Water	**Mean cal/gmole**	**Standard Deviation**
$\lambda_{12} - \lambda_{11}$	-1051.052	22.653
$\lambda_{21} - \lambda_{22}$	1802.280	55.784
	Correlation Coefficient	-0.97621976

Or writing the covariance matrix:

$$Covar\begin{pmatrix}\lambda_{12}-\lambda_{11}\\ \lambda_{21}-\lambda_{22}\end{pmatrix} = \begin{pmatrix} 513.158 & -1233.624 \\ -1233.624 & 3111.855 \end{pmatrix}$$

C-2.d. Constraint Boundary Mapping

C-2.d.i. Feed Flowrate Parameter Space

Figure C-2.d.i-1 plots the constraint boundary in the feed flowrate dimensions. Standard deviation contours for the feed-space are plotted. Boundary lines are provided for the minimum (30%) and maximum (40%) tray efficiency. Also, the boundaries are shown with the thermodynamic database at ±3 standard deviations.[1] As can be seen, the constraint boundary is sensitive to tray efficiency and insensitive to database. Also, the near-vertical direction of the constraint boundary shows that it is sensitive to acetone, but not water.

[1] Only two of the four possible boundaries (two each at minimum and maximum tray efficiencies) are shown for the thermodynamic database dimension. The database values are stepped 3 standard deviations in the top-left and bottom-right direction along their principal axis of uncertainty.

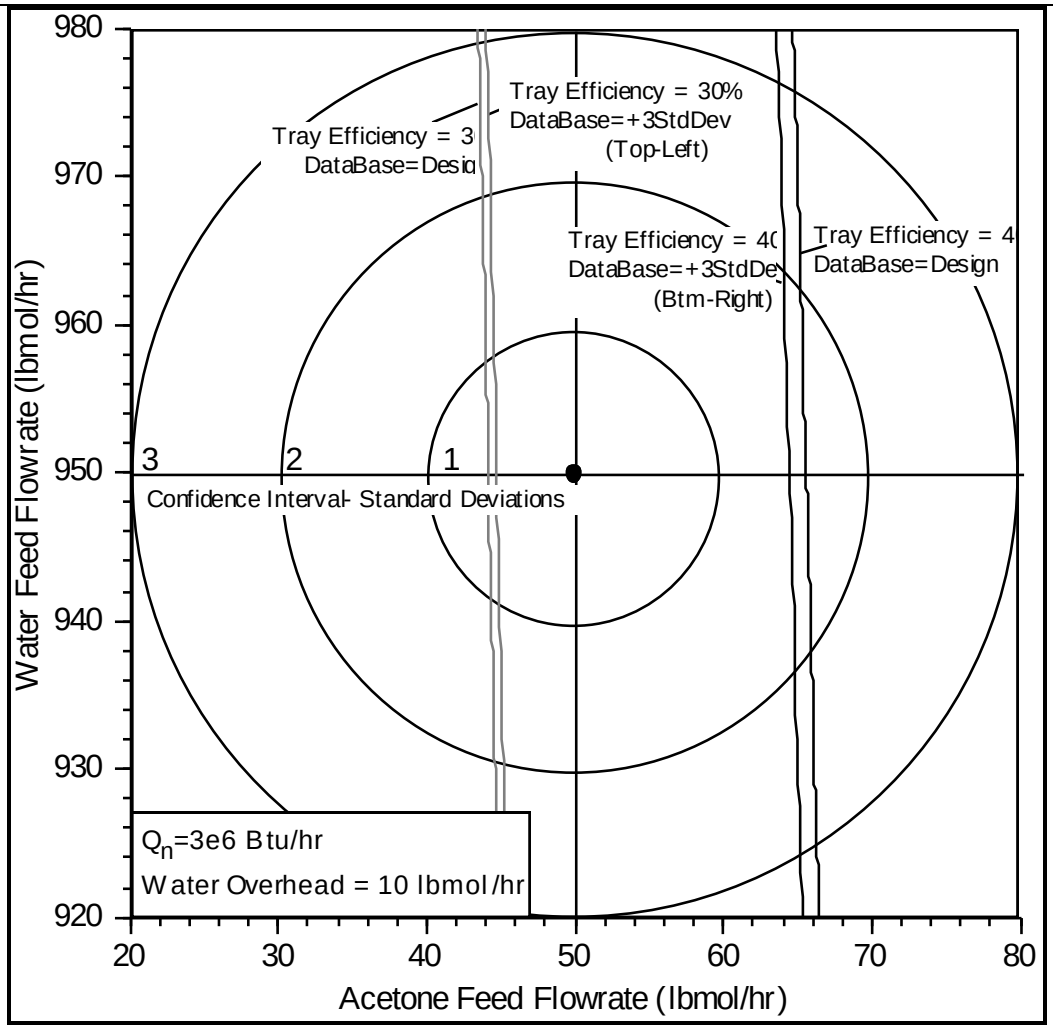


Figure C-1.d.i-1
Constraint Boundary in the Component Feed Flowrate Space, Showing Minimum and Maximum Tray Efficiency and Database Sensitivity

C-2.d.ii. Liquid Activity Coefficients Parameter Space

The constraint boundary in the database λ–space is shown in Figure C-2.d.ii-1. All other parameters are at their design value. Note that the constraint boundary is highly nonlinear, being 'banana-shaped' and open-ended in two directions. However, the 95% confidence interval for λ is tiny in comparison, suggesting that the database uncertainty is insignificant.

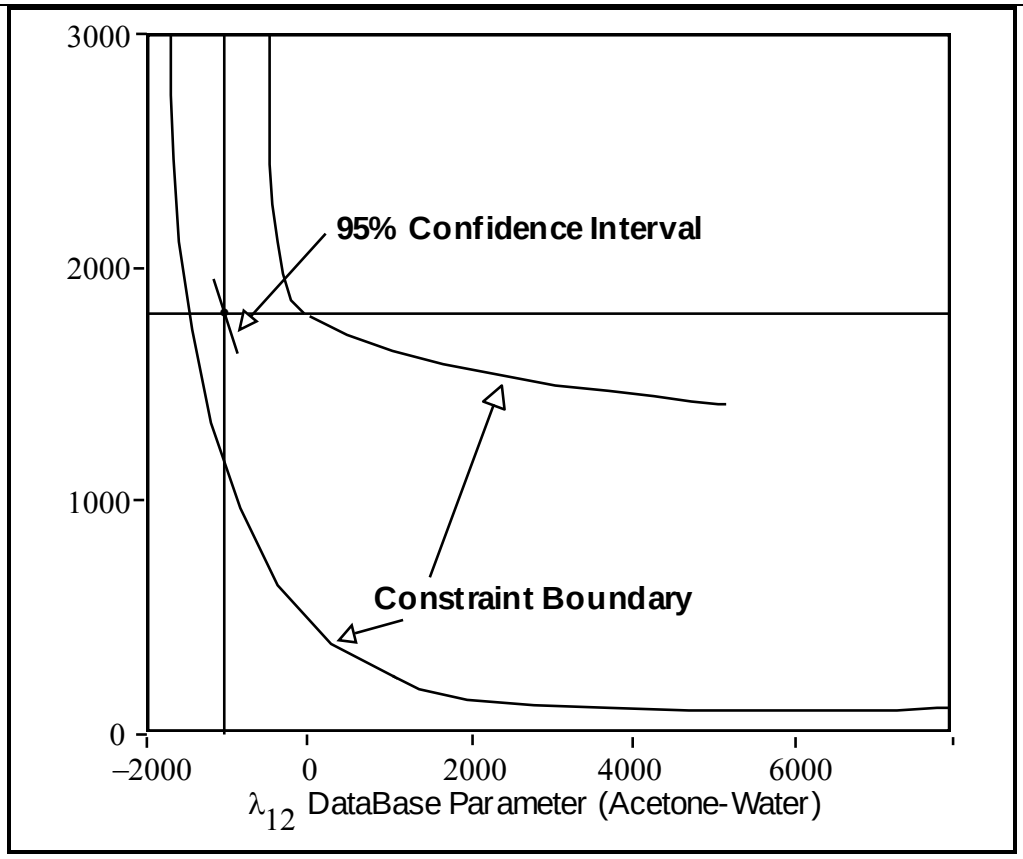


Figure C-2.d.ii-1
Map of Constraint Boundary in λ-Space

Figure C-2.d.ii-2 plots the values of the constraint boundary along the database uncertainty principal axis from -5 to +5 standard deviations. This is repeated at: (1) the design point - shifted to the origin for better comparison, (2) the design point with minimum tray efficiency, (3) the design point with maximum acetone feed flowrate, (4) the first two constraint boundary points selected by the boundary-approximation procedure. This comparison shows that the constraint boundary is fairly linear and constant with respect to the database parameters at the constraint boundary. However, the situation is very different at the design point. Thus, we could not perform a sensitivity analysis at the design point, then extrapolate this accurately to the constraint boundary.

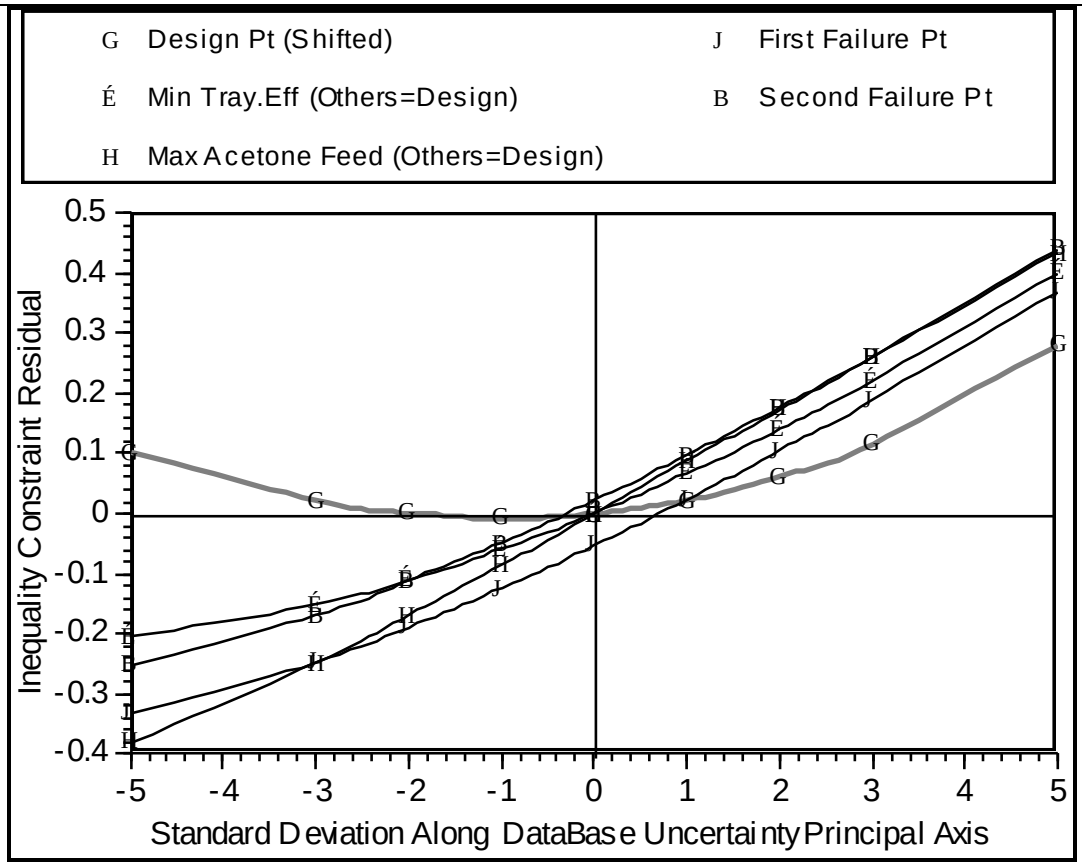


Figure C-2.d.ii-2
Shape of the Inequality Constraint as a Function of the Database Uncertainty Parameters at the Design Point and Other Constraint Boundary Points

C-3. 2-Methyl-2-Butene — Isoprene Distillation

2-Methyl-2-Butene — Isoprene (2MB1-IPM) was chosen as a narrow-boiling, multi-stage system. It has been used as a test case by Howat (1983). Relative volatilities range from 1.05 to 1.14. Database uncertainty is very significant.

C-3.a. Problem Description

Feed is nominally 50/50 lbmol/hr mixture (2MB1/IPM), with an independent, multinormal distribution of standard deviation 5/5 lbmol/hr. Tray efficiency is uniformly distributed between 90% and 100%. Isoprene minimum recovery is 95% in the bottoms. An additional specification requires at least 99% mole fraction isoprene in the bottoms.

Initially, 200 stages were specified with 100% tray efficiency and the feed to stage 125. The vapor boilup rate was 1,223 lbmol/hr. Then, I set the vapor flooding rate at 1,400 lbmol/hr (+8.1% safety factor). However, reliability was only about 62%. To increase reliability, I

increased flooding vapor flowrate and/or number of stages. While these were exploratory studies (reliability estimate is approximate), the following information was obtained:

Table C-3.a-1
Sensitivity of Design Reliability to Number of Stages and Vapor Rate for the 2MB1-IPM Test Problem

Number of Stages	Vapor Rate Safety Factor (Basis: Design; 100% Eff)	Approximate Reliability (%)
200	21%	62
200	38%	69
200	60%	73
250	10%	73
250	60%	80

A 60% vapor flowrate safety factor only increased the reliability to 73%. The same increase was noted by increasing the number of stages by 50. Such large vapor rate safety factors would be very expensive and increasing the number of stages is probably a better alternative. Whatever the case, these reliabilities are unacceptably low for such an expensive column, so further sensitivity analysis and a more accurate statistical analysis of the thermodynamic data is warranted.

For this work I selected 200 stages to minimize the computational effort. 21% safety factor was used for the vapor boilup, which serves as a test for low reliability problems.

The following section lists the column profile information for the base case design. The design specifications and uncertainties are listed in the following tables.

Table C-3.a-2
2MB1-IPM Test Problem Inequality Constraints

CV-1) Reboiler Vapor Rate ≤ 1,477.08 lbmol/hr

CV-2) IPM Bottoms Recovery ≥ 95%

A) IPM Bottoms Purity ≥ 99 mol%

Table C-3.a-3
2MB1-IPM Test Problem Feed Flowrate and Tray Efficiency Uncertainties

- 2MB1 feed flow is normally distributed with a mean of 50 lbmol/hr and a standard deviation of 5 lbmol/hr.
- IPM feed flow is normally distributed with a mean of 50 lbmol/hr and a standard deviation of 5 lbmol/hr.
- Tray efficiency is uniformly distributed with a minimum of 90% and a maximum of 100%.

C-3.b. Converged Run listing

```
--------------------------------------------------------------------------------
Number of Stages =  200
Total  Condenser   - Partial Reboiler
Feed stream feed              to stage  125
Draw stream ovd               from stage  1
Draw stream btms               from stage  200
Murphree vapor efficiency =  100.0 %
Condenser Pressure =  24.65 psia
Vapor Ovd Pressure =  27.65 psia
Reboiler  Pressure =  47.45 psia
Top stage:  Recovery =  5.0000E-02 for component  2
Bottom stage:  Component mole fraction = 1.0000E-02 for component 2
Convergence criterion        =  5.0000E-06
Convergence option   =   fractional percentage

--------------------------------------------------------------------------------
Feed Streams for Operation

feed              to stage  125
Temp (deg F):  153.77530     MolWt:         69.12700
Press (psia):   35.00000     Spec:         P,Tbub
Quality:         .00     H (Btu/lbmol):  67106.0500

Component          LbMoleFlow     Mole %   LbMassFlow     Mass %
2-methyl-butene-1     50.00000      50.0000   3506.750      50.7291
isoprene              50.00000      50.0000   3405.950      49.2709
Total                100.0000      100.0000   6912.700     100.0000
```

Draw Streams for Operation

ovd from stage 1
Temp (deg F): 116.51000 MolWt: 70.03812
Press (psia): 24.65000 Spec: None
Quality: .00 H (Btu/lbmol): 66700.2600

Component	LbMoleFlow	Mole %	LbMassFlow	Mass %
2-methyl-butene-1	49.52020	95.1942	3473.099	95.3259
isoprene	2.500000	4.8058	170.2975	4.6741
Total	52.02020	100.0000	3643.397	100.0000

btms from stage 200
Temp (deg F): 163.48150 MolWt: 68.13916
Press (psia): 47.44978 Spec: None
Quality: .00 H (Btu/lbmol): 66250.2000

Component	LbMoleFlow	Mole %	LbMassFlow	Mass %
2-methyl-butene-1	.4798009	1.0000	33.65083	1.0293
isoprene	47.50000	99.0000	3235.653	98.9707
Total	47.97980	100.0000	3269.303	100.0000

Stages for Operation

Stg	Press	Temp	Vap-lbm	Liq-lbm	Feed	LiqDraw	VapDraw	Heat-Btu/hr
1	24.65	116.51	0.0	77086.1	.0	3643.4	.0	-.1403E+08
2	27.65	123.28	80729.5	79069.2	.0	.0	.0	0.0000E+00
15	28.95	126.15	82876.5	79246.0	.0	.0	.0	0.0000E+00
30	30.45	129.37	83055.2	79422.6	.0	.0	.0	0.0000E+00
45	31.95	132.51	83200.8	79565.9	.0	.0	.0	0.0000E+00
60	33.45	135.60	83311.0	79673.8	.0	.0	.0	0.0000E+00
75	34.95	138.63	83389.3	79750.3	.0	.0	.0	0.0000E+00
90	36.45	141.62	83448.9	79809.4	.0	.0	.0	0.0000E+00
105	37.95	144.53	83513.8	79875.6	.0	.0	.0	0.0000E+00
125	39.95	148.25	83652.4	86774.5	6912.7	.0	.0	0.0000E+00
140	41.45	151.07	83604.4	86872.4	.0	.0	.0	0.0000E+00
155	42.95	154.28	83455.3	86706.0	.0	.0	.0	0.0000E+00
170	44.45	157.81	83175.4	86431.4	.0	.0	.0	0.0000E+00
185	45.95	160.94	83122.9	86399.4	.0	.0	.0	0.0000E+00
200	47.45	163.48	83318.5	0.0	.0	3269.3	.0	0.1396E+08

Stage Composition and Relative Volatilities for Operation

Stg	2MB1 Liquid Mole %	K-value 2MB1	K-value Isoprene	Relative Volatility
1	95.19%	1.003	.9496	1.056
2	94.94%	1.003	.9505	1.055
15	91.51%	1.005	.9498	1.058
30	87.08%	1.008	.9493	1.061
45	82.01%	1.011	.9489	1.066
60	76.24%	1.016	.9490	1.071
75	69.85%	1.022	.9496	1.076
90	63.19%	1.029	.9509	1.082
105	56.84%	1.036	.9528	1.087
125	49.97%	1.044	.9556	1.093
140	43.50%	1.053	.9589	1.098
155	30.66%	1.074	.9674	1.110
170	14.73%	1.104	.9820	1.124
185	4.67%	1.126	.9938	1.133

200	1.00%	1.135	.9986	1.136

C-3.c. Database Regression

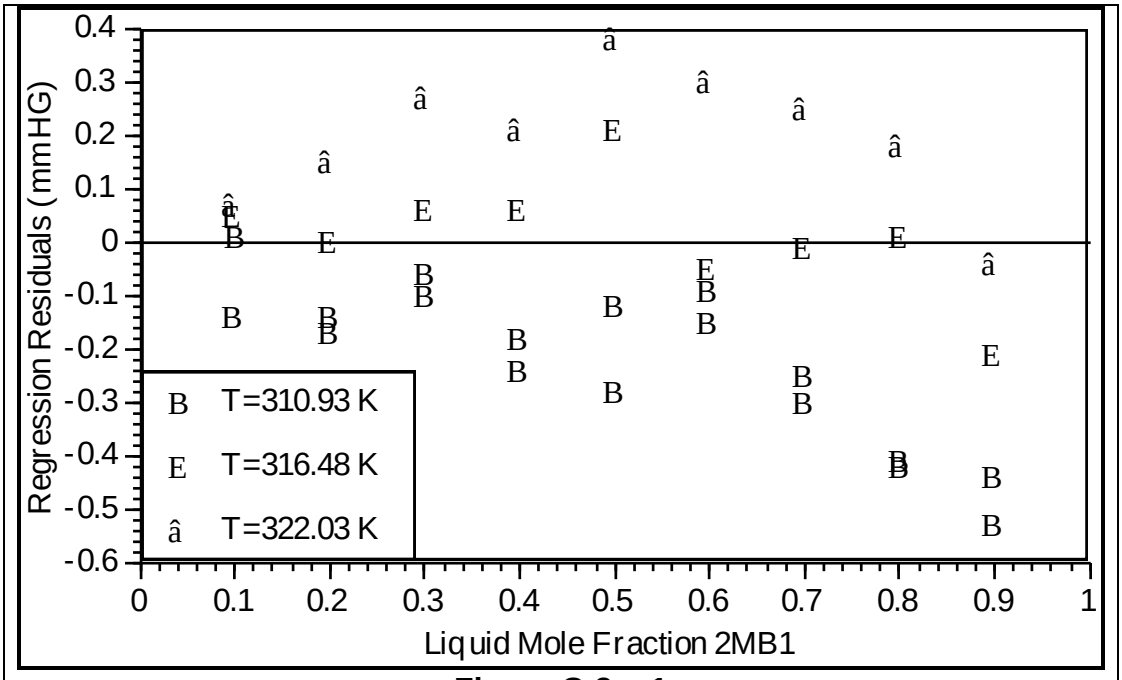


Figure C-3.c-1
Residual Pressure Plot of Modified Wilson Regression of Shanker et al. (1981) 2MB1-IPM VLE Data

The Modified Wilson coefficients and their uncertainties used for this work are shown in Table C-3.c-1:

Table C-3.c-1
2MB1-IPM Test Problem: Statistical Distribution for Modified Wilson Solution Model Coefficients

	Bivariate Normal Distribution	
1-2MB1 2-IPM	**Mean cal/gmole**	**Standard Deviation**
$\lambda_{12} - \lambda_{11}$	-50.33145	149.738
$\lambda_{21} - \lambda_{22}$	76.894	153.02
	Correlation Coefficient	-0.9999655193

Or writing the covariance matrix:

$$Covar\begin{pmatrix}\lambda_{12}-\lambda_{11}\\ \lambda_{21}-\lambda_{22}\end{pmatrix} = \begin{pmatrix}22421.47 & -22912.12\\ -22912.12 & 23415.12\end{pmatrix}$$

C-3.d. Constraint Boundary Mapping

C-3.d.i. Feed Flowrate Parameter Space

Figure C-3.d.i-1 plots the constraint boundary in the feed flowrate dimensions. Standard deviation contours for the feed-space are provided. Note that the constraint boundary passes well inside these contours, which confirms the fact that the design reliability is very low.

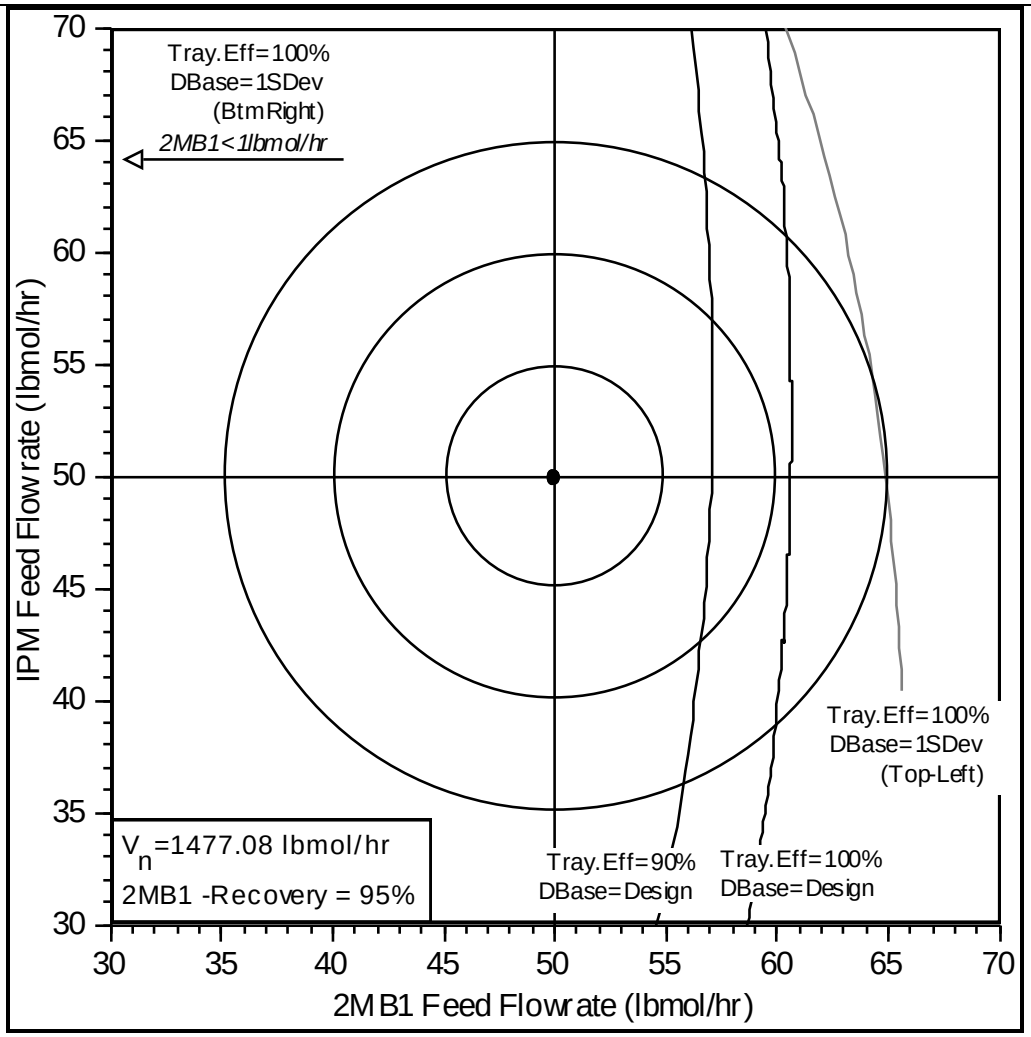


Figure C-3.d.i-1
Constraint Boundary in the Component Feed Flowrate Space, Showing Minimum and Maximum Tray Efficiency and Database Sensitivity

Contours are provided for the minimum (90%) and maximum (100%) tray efficiency. Tray efficiency is a significant parameter. Also, the thermodynamic database was set at ±1

standard deviation with a100% tray efficiency. In the top-left principal axis direction, the boundary is expanded and its shape is altered. In the bottom-right principal axis direction, the design fails for all feeds. (Actually, a minimum flowrate boundary exists for 2MB1 at about 1 lbmol/hr, which is statistically insignificant.) Consequently, it is clear that database is a very significant uncertainty.

C-3.d.ii. Liquid Activity Coefficients Parameter Space

The constraint boundary in the database λ–space is shown in Figure C-3.d.ii-1. All other parameters are at their design value. Note that the constraint boundary is a skinny 'banana-shaped' diagonal ellipse. Also, note that the 95% confidence interval falls outside the boundary.

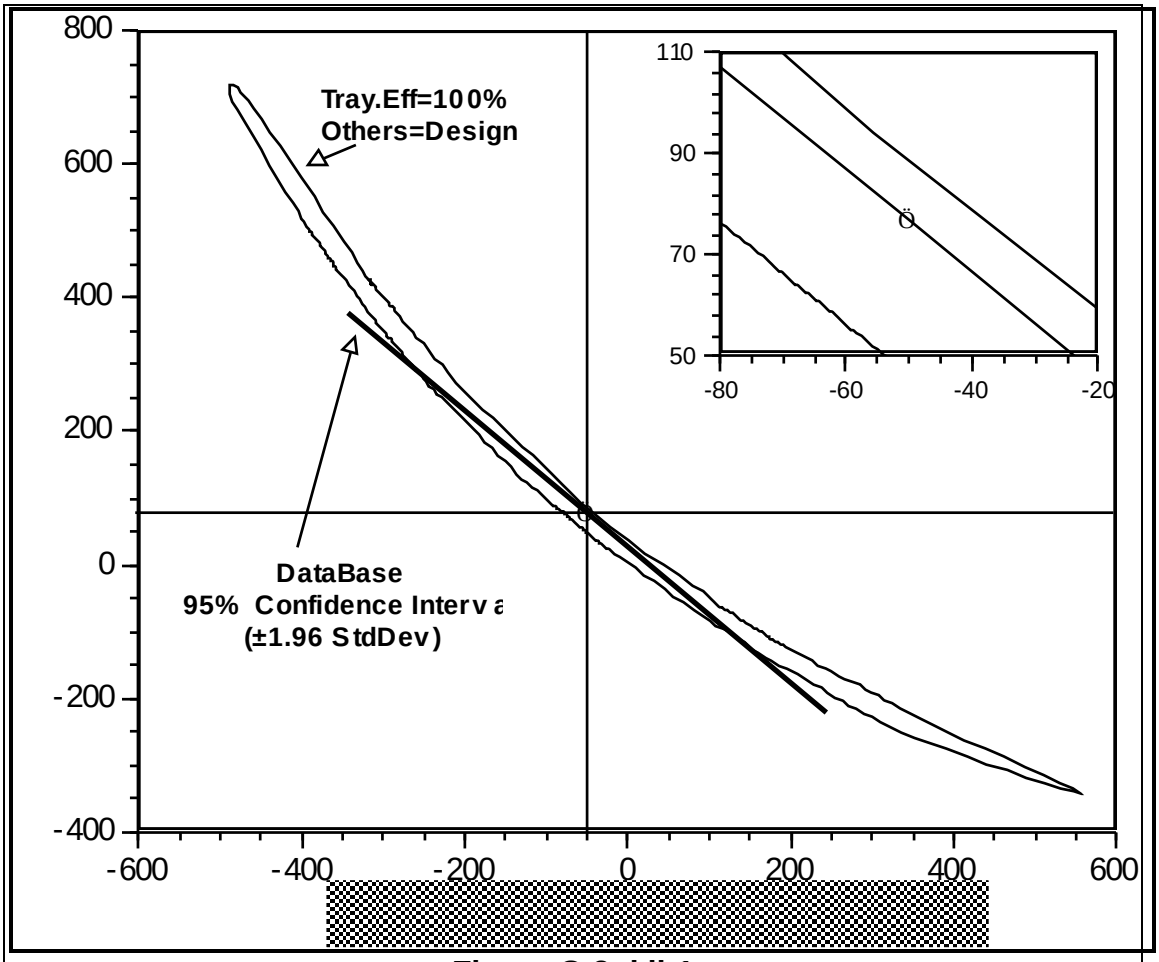


Figure C-3.d.ii-1
Map of Constraint Boundary in λ–Space

Figure C-3.d.ii-2 plots the values of the constraint boundary along the database uncertainty principal axis from -3 to +3 standard deviations. This is repeated at: (1) the design

point, (2) the design point with minimum tray efficiency, (3) the maximum 2MB1 feed flowrate, (4) +3 standard deviations of 2MB1 feed flowrate, and (5) the first and second boundary point selected by the convergence procedure. Here, the relationship is fairly symmetrical, but nonconstant and nonideal over the region.

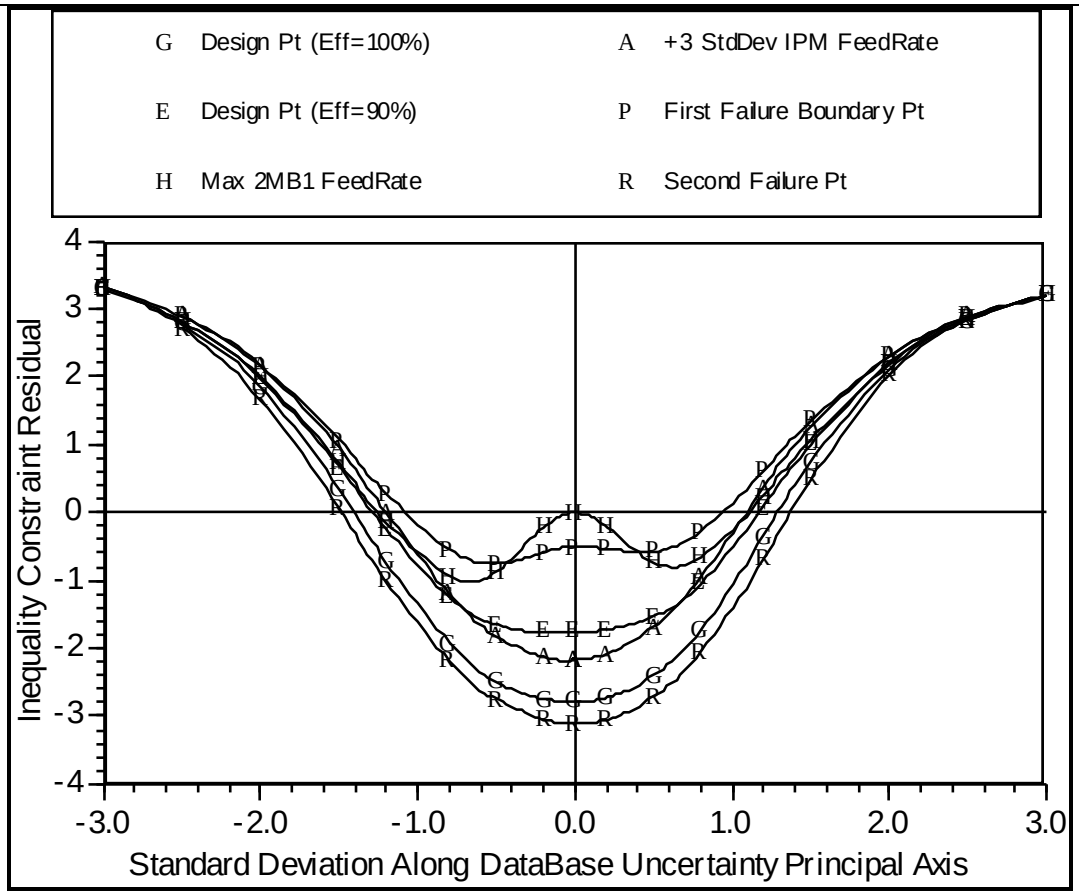


Figure C-3.d.ii-2
Shape of the Inequality Constraint as a Function of the Database Uncertainty Parameters at the Design Point and Other Constraint Boundary Points

C-4. Acetone-Benzene-Toluene Distillation

Toluene was chosen as a third component, because binary vapor-liquid equilibria (VLE) data exist for Acetone-Toluene and Benzene-Toluene. Ternary VLE data exist but were not used. Toluene is the heaviest component, so it could simply be added to the Acetone-Benzene test problem in section C-1, without major changes.

C-4.a. Problem Description

As with the acetone-benzene test case, the column has 25 stages (including condenser) with the feed to stage 15, and a total condenser operating at atmospheric pressure. Vapor overhead pressure drop is 3 psi. Pressure drop per stage is 0.1 psi/stage.

Feed flows were changed slightly by decreasing the benzene and including toluene. The nominal flowrate is 500/375/125 lbmol/hr (acetone/benzene/toluene). Feed is a saturated liquid at 20 psia. Feed uncertainties were decreased in order to keep the design reliability approximately equal to the acetone-benzene test case.

The same process specifications were used as for the acetone-benzene test problem, except for specifying a 10% vapor boilup safety factor. This gives:

Table C-4.a-1
Acetone-Benzene-Toluene Test Problem Inequality Constraints

CV-1)	Reboiler Vapor Rate ≤ 97,684 lbm/hr
CV-2)	Acetone Overhead Recovery ≥ 80%
A)	Acetone Overhead Purity ≥ 95 mol%
B)	Acetone Bottoms Impurity ≤ 25 mol%

The following statistical uncertainties were used for feed flowrate and tray efficiency:

Table C-4.a-2
Acetone-Benzene-Toluene Test Problem Feed Flowrate and Tray Efficiency Uncertainties

- Acetone feed flow is normally distributed with a mean of 500 lbmol/hr and a standard deviation of 12 lbmol/hr.
- Benzene feed flow is normally distributed with a mean of 375 lbmol/hr and a standard deviation of 25 lbmol/hr.
- Toluene feed flow is normally distributed with a mean of 125 lbmol/hr and a standard deviation of 30 lbmol/hr.
- Tray efficiency is uniformly distributed with a minimum of 50% and a maximum of 70%.

C-4.b. Converged Run listing

Number of Stages = 25
Total Condenser - Partial Reboiler
Feed stream feed to stage 15
Draw stream ovd from stage 1
Draw stream btms from stage 25
Murphree vapor efficiency = 60.00 %
Condenser Pressure = 15.00 psia
Vapor Ovd Pressure = 18.00 psia
Reboiler Pressure = 20.30 psia
Top stage: Recovery = .8000 for component 1
Top stage 2: Component mole fraction = 0.95 for component 1
Convergence criterion = 5.0000E-07
Convergence option = fractional percentage

Feed Streams for Operation

feed to stage 15
Temp (deg F): 164.77470 MolWt: 69.50010
Press (psia): 20.00000 Spec: P,Tbub
Quality: .00 H (Btu/lbmol): -4213.3020

Component	LbMoleFlow	Mole %	LbMassFlow	Mass %
Acetone	500.0000	50.0000	29040.00	41.7841
Benzene	375.0000	37.5000	29293.12	42.1483
Toluene	125.0000	12.5000	11517.63	16.5721
Total	1000.000	100.0000	69500.10	100.0000

Draw Streams for Operation

ovd from stage 1
Temp (deg F): 135.00290 MolWt: 59.08214
Press (psia): 15.00000 Spec: None
Quality: .00 H (Btu/lbmol): -6131.0170

Component	LbMoleFlow	Mole %	LbMassFlow	Mass %
Acetone	400.0000	95.0000	23232.00	93.3886
Benzene	21.04074	4.9972	1643.598	6.6070
Toluene	0.1181524E-01	.0028	1.088668	.0044
Total	421.0526	100.0000	24876.69	100.0000

btms from stage 25
Temp (deg F): 185.16160 MolWt: 77.68246
Press (psia): 20.29999 Spec: None
Quality: .00 H (Btu/lbmol): -2332.5940

Component	LbMoleFlow	Mole %	LbMassFlow	Mass %
Acetone	100.0000	17.2727	5808.000	12.9141
Benzene	353.9593	61.1384	27649.53	61.4788
Toluene	124.9882	21.5889	11516.54	25.6071
Total	578.9474	100.0000	44974.07	100.0000

Stages for Operation

Stg	Press	Temp	Vap-lbm	Liq-lbm	Feed	LiqDraw	VapDraw	Heat-Btu/hr
1	15.00	135.00	0.0	51421.9	.0	24876.7	.0	0.5107E+06
2	18.00	145.21	76298.6	52878.2	.0	.0	.0	0.0000E+00
3	18.10	145.90	77754.8	53110.4	.0	.0	.0	0.0000E+00
4	18.20	146.61	77987.1	53352.3	.0	.0	.0	0.0000E+00
5	18.30	147.34	78229.0	53604.2	.0	.0	.0	0.0000E+00
6	18.40	148.10	78480.9	53866.8	.0	.0	.0	0.0000E+00
7	18.50	148.89	78743.5	54140.9	.0	.0	.0	0.0000E+00
8	18.60	149.72	79017.6	54427.6	.0	.0	.0	0.0000E+00
9	18.70	150.60	79304.3	54728.4	.0	.0	.0	0.0000E+00
10	18.80	151.54	79605.1	55045.7	.0	.0	.0	0.0000E+00
11	18.90	152.55	79922.4	55383.1	.0	.0	.0	0.0000E+00
12	19.00	153.68	80259.8	55746.2	.0	.0	.0	0.0000E+00
13	19.10	154.96	80622.8	56142.4	.0	.0	.0	0.0000E+00
14	19.20	156.48	81019.1	56340.9	.0	.0	.0	0.0000E+00
15	19.30	160.18	81217.6	127295.0	69850.8	.0	.0	0.0000E+00
16	19.40	160.61	82320.9	127471.5	.0	.0	.0	0.0000E+00
17	19.50	161.10	82497.4	127703.2	.0	.0	.0	0.0000E+00
18	19.60	161.67	82729.2	128015.2	.0	.0	.0	0.0000E+00
19	19.70	162.38	83041.1	128443.4	.0	.0	.0	0.0000E+00
20	19.80	163.30	83469.3	129039.4	.0	.0	.0	0.0000E+00
21	19.90	164.52	84065.3	129875.6	.0	.0	.0	0.0000E+00
22	20.00	166.24	84901.5	131050.0	.0	.0	.0	0.0000E+00
23	20.10	168.73	86076.0	132681.7	.0	.0	.0	0.0000E+00
24	20.20	172.46	87707.7	133811.8	.0	.0	.0	0.0000E+00
25	20.30	185.16	88837.7	0.0	.0	44974.1	.0	-.5124E+06

Stage Composition and Relative Volatilities for Operation

Stg	Acetone Liq. Mol%	Benzene Liq. Mol%	K-value Acetone	K-value Benzene	K-value Toluene	Relative Volatility Acet-Benz	Relative Volatility Tol-Benz
1	95.00%	5.00%	1.020	.6275	.2325	1.625	.3705
2	93.31%	6.69%	1.026	.6327	.2395	1.622	.3785
3	91.53%	8.46%	1.034	.6323	.2390	1.635	.3779
4	89.68%	10.30%	1.042	.6319	.2384	1.650	.3773
5	87.74%	12.23%	1.052	.6316	.2379	1.665	.3766
6	85.70%	14.24%	1.062	.6314	.2374	1.681	.3760
7	83.57%	16.34%	1.073	.6313	.2369	1.700	.3753
8	81.33%	18.52%	1.085	.6314	.2366	1.719	.3747
9	78.96%	20.79%	1.099	.6316	.2363	1.741	.3741
10	76.46%	23.13%	1.115	.6322	.2361	1.764	.3735
11	73.79%	25.53%	1.134	.6331	.2362	1.791	.3730
12	70.90%	27.98%	1.156	.6347	.2365	1.821	.3726
13	67.74%	30.44%	1.184	.6373	.2374	1.857	.3725
14	64.16%	32.85%	1.219	.6416	.2392	1.900	.3728
15	56.40%	35.53%	1.322	.6590	.2473	2.007	.3753
16	56.00%	35.93%	1.327	.6596	.2477	2.012	.3755
17	55.40%	36.50%	1.334	.6605	.2481	2.019	.3757
18	54.54%	37.32%	1.344	.6618	.2487	2.030	.3758
19	53.30%	38.49%	1.359	.6637	.2495	2.047	.3759
20	51.50%	40.15%	1.381	.6667	.2507	2.072	.3760
21	48.91%	42.46%	1.417	.6716	.2526	2.110	.3761
22	45.24%	45.64%	1.473	.6798	.2559	2.167	.3764
23	40.12%	49.85%	1.564	.6943	.2619	2.253	.3772

24	33.20%	55.06%	1.719	.7213	.2734	2.383	.3790
25	17.27%	61.14%	2.349	.8545	.3327	2.749	.3893

C-4.c. Database Regression

The three binary pairs were regressed individually to obtain binary coefficients for the liquid solution model. The solution model was trusted to extrapolate into the ternary region. Acetone-benzene binary coefficients are the same as in section C-1.c. Two data sets are used for the Acetone-Toluene binary set and the pressure residual error plot is shown in Figure C-4.c-1. The Rosanoff et al. (1914) set appears to be inferior and would be eliminated in a rigorous design study. However, for this test I was interested in a database with large uncertainty, so this set was kept.

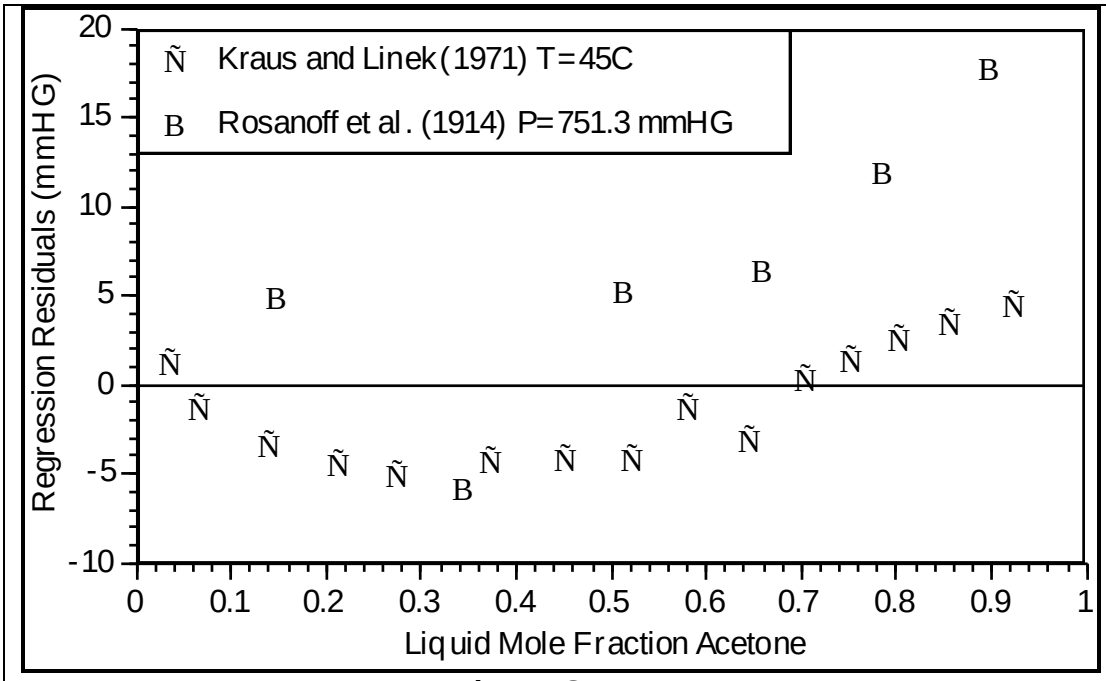


Figure C-4.c-1
Residual Pressure Plot of Modified Wilson Regression of Acetone-Toluene VLE Data

Two TPxy data sets were available at 760 mmHG for benzene-toluene. These were regressed and the pressure residual error plots are shown in Figure C-4.c-2. Obviously, the Rollet et al. (1956) data set is much more precise, however, both data sets passed the thermodynamic consistency tests of Gmehling and Onken (1977ff).

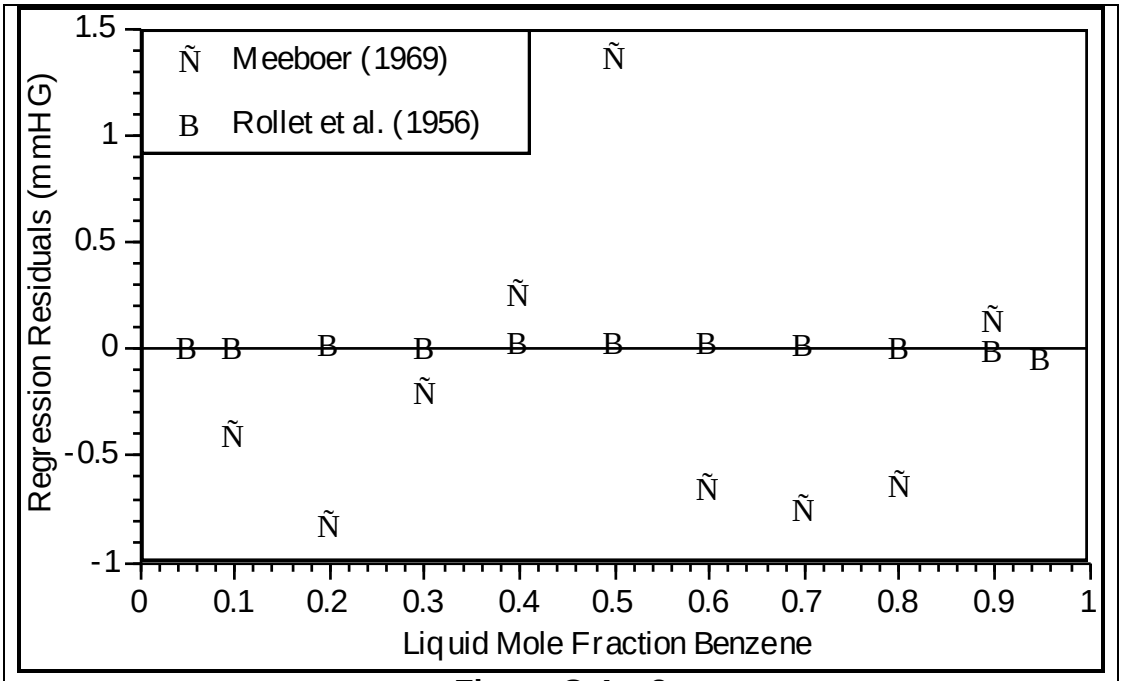


Figure C-4.c-2
Residual Pressure Plot of Modified Wilson Regression of Benzene-Toluene VLE Data

The bivariate normal distributions for the two binary sets are shown in Tables C-4.c-1 and C-4.c-2.

Table C-4.c-1
Acetone – Toluene: Statistical Distribution for Modified Wilson Solution Model Coefficients

	Bivariate Normal Distribution	
1-Acetone 2-Toluene	**Mean cal/gmole**	**Standard Deviation**
$\lambda_{12} - \lambda_{11}$	751.381	78.668
$\lambda_{21} - \lambda_{22}$	39.5464	52.120
	Correlation Coefficient	-0.9889536

Or writing the covariance matrix:

$$Covar\begin{pmatrix}\lambda_{12}-\lambda_{11}\\ \lambda_{21}-\lambda_{22}\end{pmatrix} = \begin{pmatrix} 6188.654 & -4054.884 \\ -4054.884 & 986.525 \end{pmatrix}$$

Table C-4.c-2
Benzene – Toluene: Statistical Distribution for Modified Wilson Solution Model Coefficients

1-Benzene 2-Toluene	Bivariate Normal Distribution	
	Mean cal/gmole	Standard Deviation
$\lambda_{12} - \lambda_{11}$	-200.517	28.079
$\lambda_{21} - \lambda_{22}$	270.079	45.828
	Correlation Coefficient	-0.9998827

Or writing the covariance matrix:

$$Covar\begin{pmatrix}\lambda_{12}-\lambda_{11}\\ \lambda_{21}-\lambda_{22}\end{pmatrix} = \begin{pmatrix} 788.430 & -1286.654 \\ -1286.654 & 2100.206 \end{pmatrix}$$

C-4.d. Constraint Boundary Mapping

C-4.d.i. Feed Flowrate Parameter Space

Figure C-4.d.i-1 plots the constraint boundary in the feed flowrate dimensions of acetone and benzene. As a third dimension, toluene is shown at -3, 0, +3 feed flowrate standard deviations with separate boundary lines. The boundary is also shown for the minimum tray efficiency.

As toluene feed increases, the acetone overhead purity specification is harder to maintain. However, the benzene bottoms impurity is easier to maintain. However, note that this effect is small in comparison with the feed flowrate uncertainty. This is seen by the comparison between the size of the 1,2,3 standard deviation of the feed flowrate versus the larger distances of the constraint boundary. Tray efficiency is a much more significant uncertainty, and the design actually fails at design conditions with the minimum tray efficiency of 50% (apparently not a very good design).

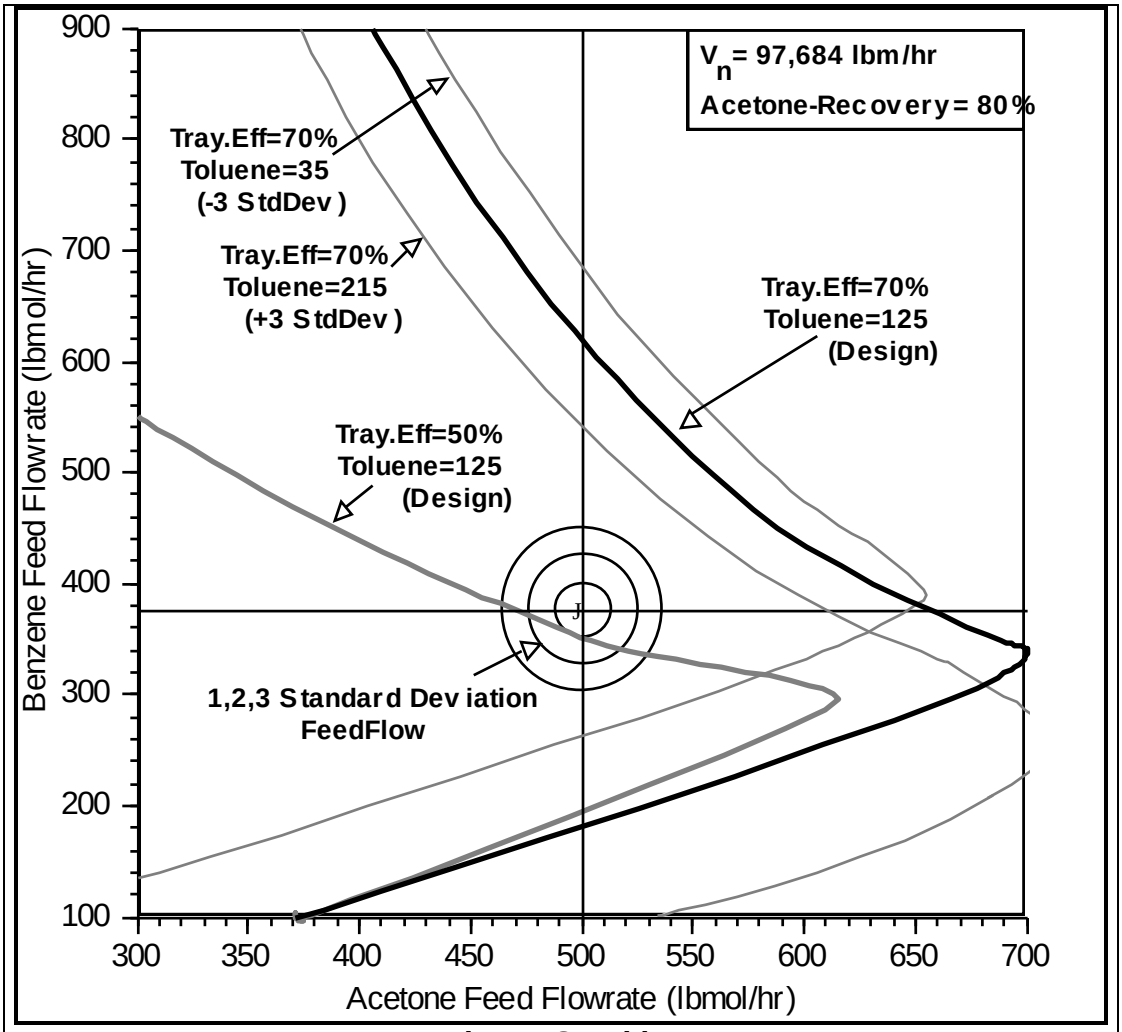


Figure C-4.d.i-1
Constraint Boundary in the Component Feed Flowrate Space, Showing Minimum and Maximum Tray Efficiency Sensitivity

C-4.d.ii. Liquid Activity Coefficients Parameter Space

For a ternary system, the Modified Wilson solution model (Tsuboka and Katayama, 1975) has three binary pair parameters: (1) Acetone-Benzene, (2) Acetone-Toluene, and (3) Benzene-Toluene. The constraint boundaries were mapped in these three parameter spaces and are shown in Figures C-4.d.ii-1,a,b,c. For these maps all other parameters, including the other two Modified Wilson coefficient pairs, are set at their design values. Sensitivity analysis showed that the three binary pairs are practically independent. That is, the constraint boundaries are unaffected by the values of the other component binary pair coefficient values. This is because the constraint boundary is so large in comparison to the statistical confidence intervals of these parameters.

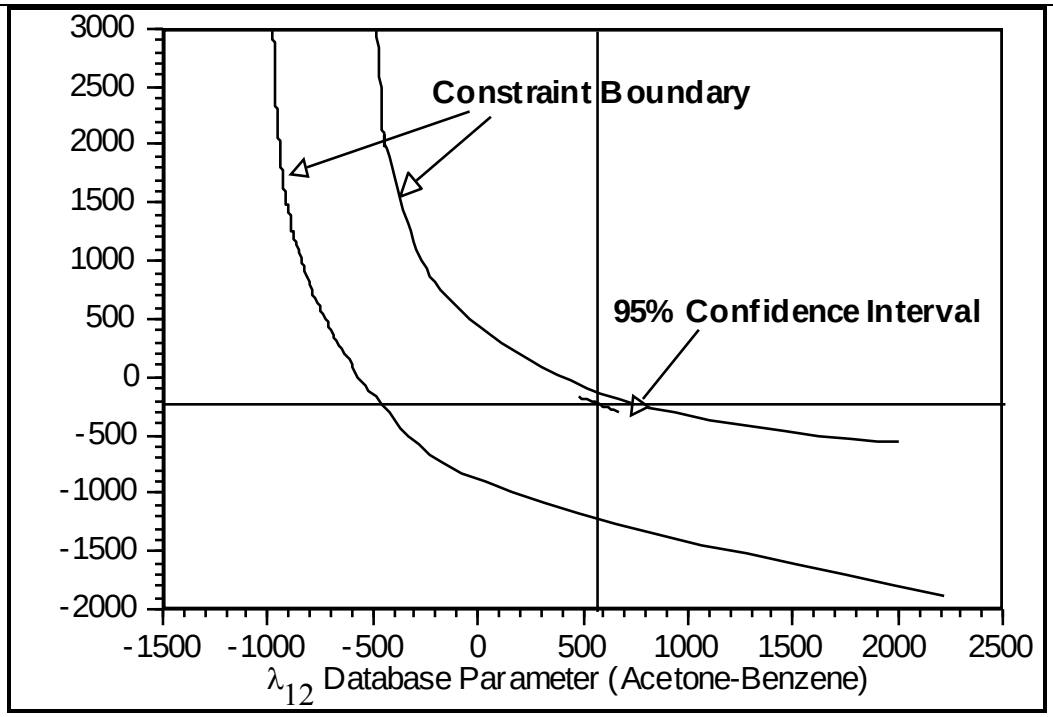


Figure C-4.d.ii-1a
Map of Constraint Boundary in Acetone-Benzene λ-Space

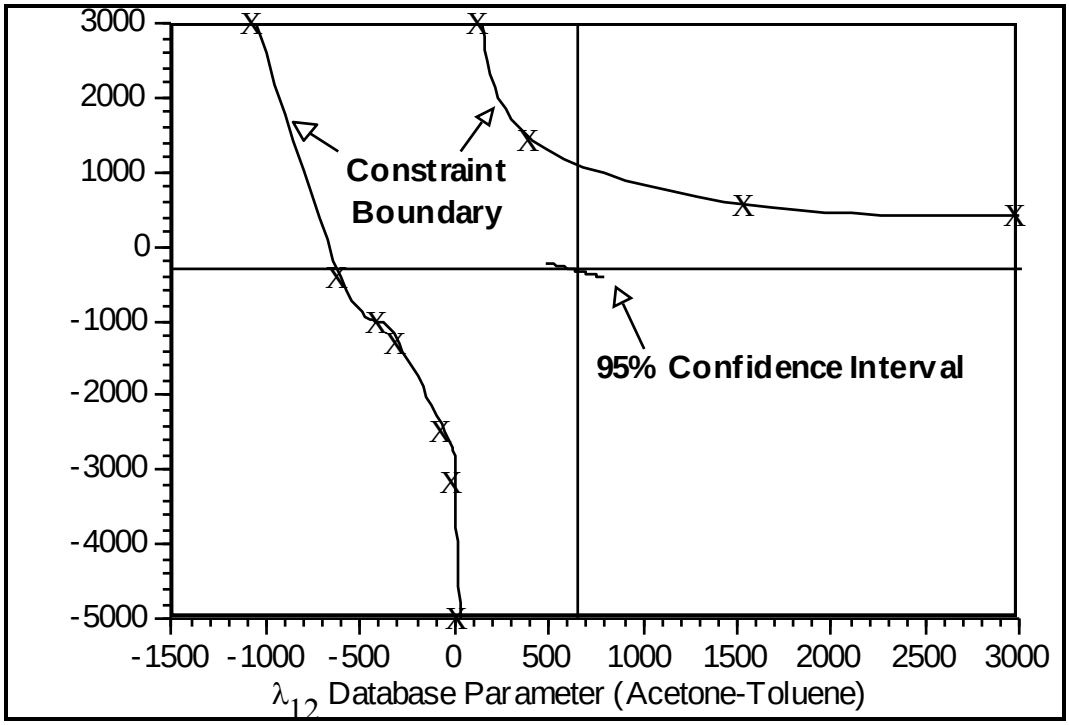


Figure C-4.d.ii-1b
Map of Constraint Boundary in Acetone-Toluene λ-Space

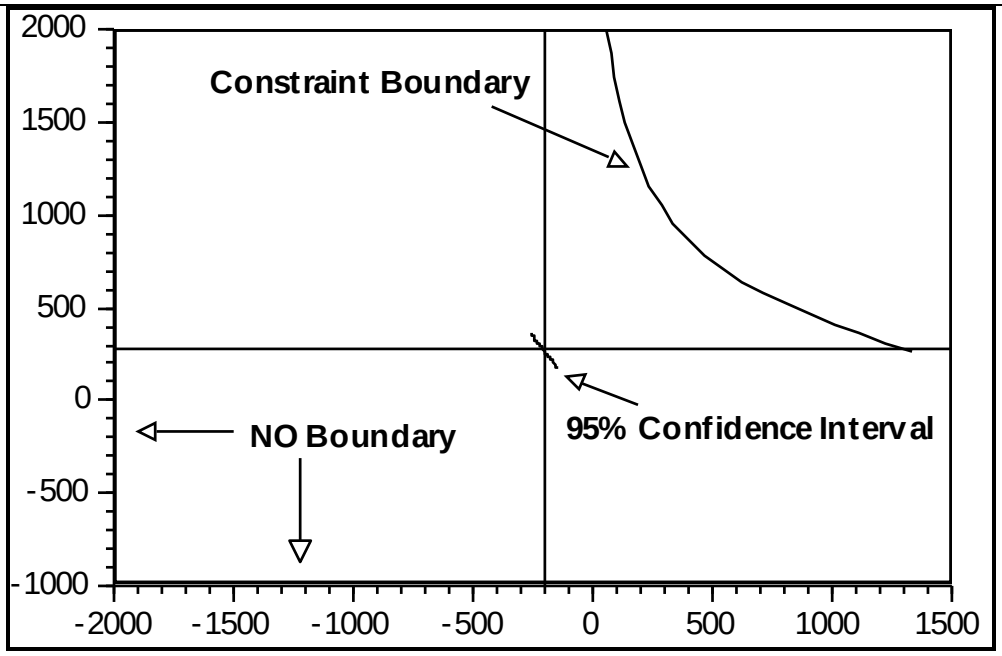


Figure C-4.d.ii-1c
Map of Constraint Boundary in Benzene-Toluene λ-Space

Figures C-4.d.ii-2 a,b,c plot the values of the inequality constraint residuals along the database principal axis of the acetone-benzene thermodynamic coefficients. This is done at -3, 0, +3 standard deviations along the principal axis of the acetone-toluene and benzene-toluene thermodynamic coefficients to give nine slopes for comparison at each point. Each graph compares the slopes for a single point. Six points are checked: (1) design point, (2) maximum acetone feed axial boundary point, (3) maximum benzene feed axial boundary point, (4) minimum benzene feed axial boundary point, (5) design point with minimum tray efficiency boundary point. In all cases, the slope is quadratic and a linear approximation is inaccurate. However, fitting a quadratic curve at the design point and extrapolating throughout the region is not accurate either, because the slope changes.

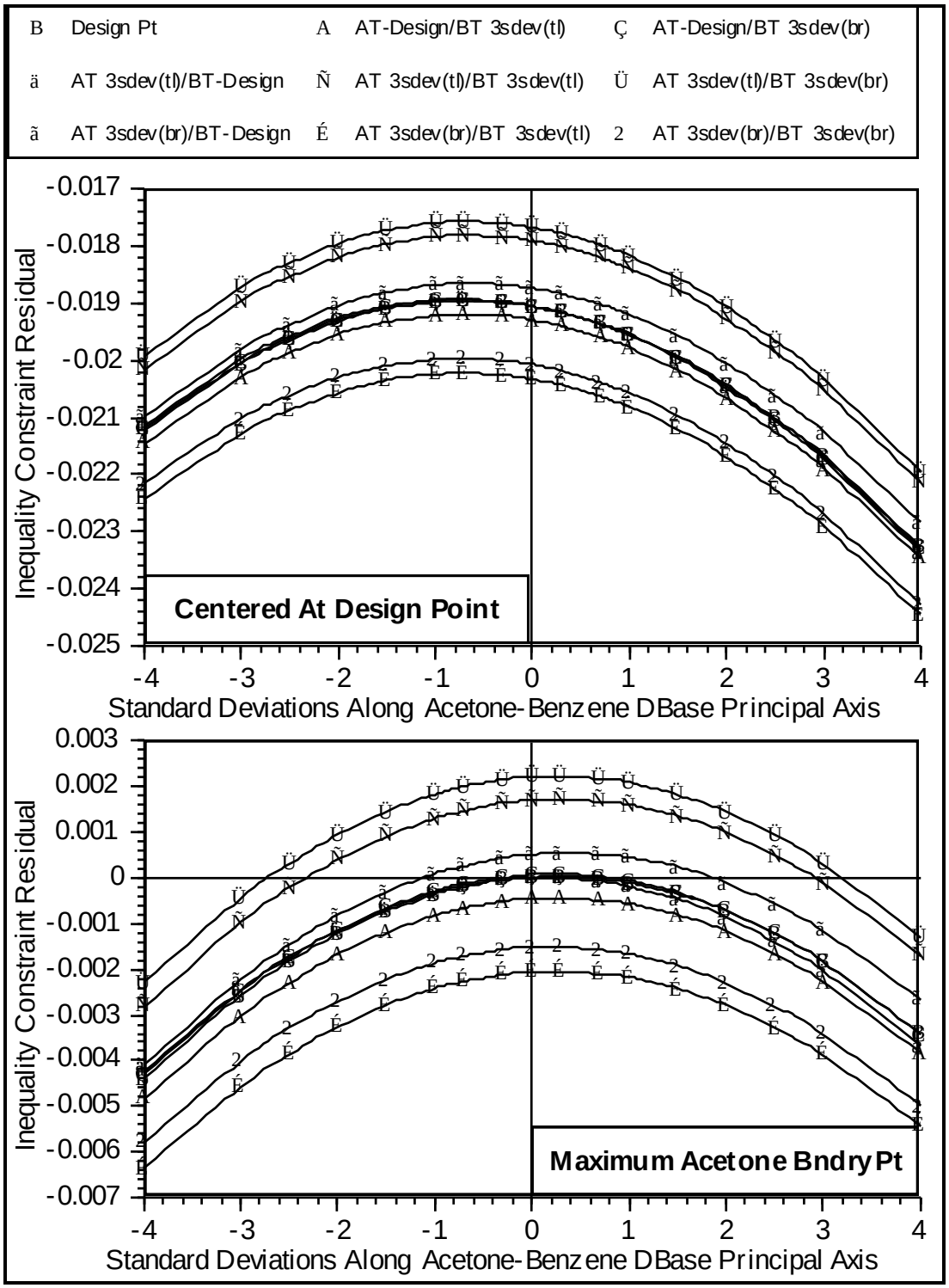


Figure C-4.d.ii-2 a
Shape of Inequality Constraint as a Function of the Database Uncertainty Parameters Along Their Principal Axes

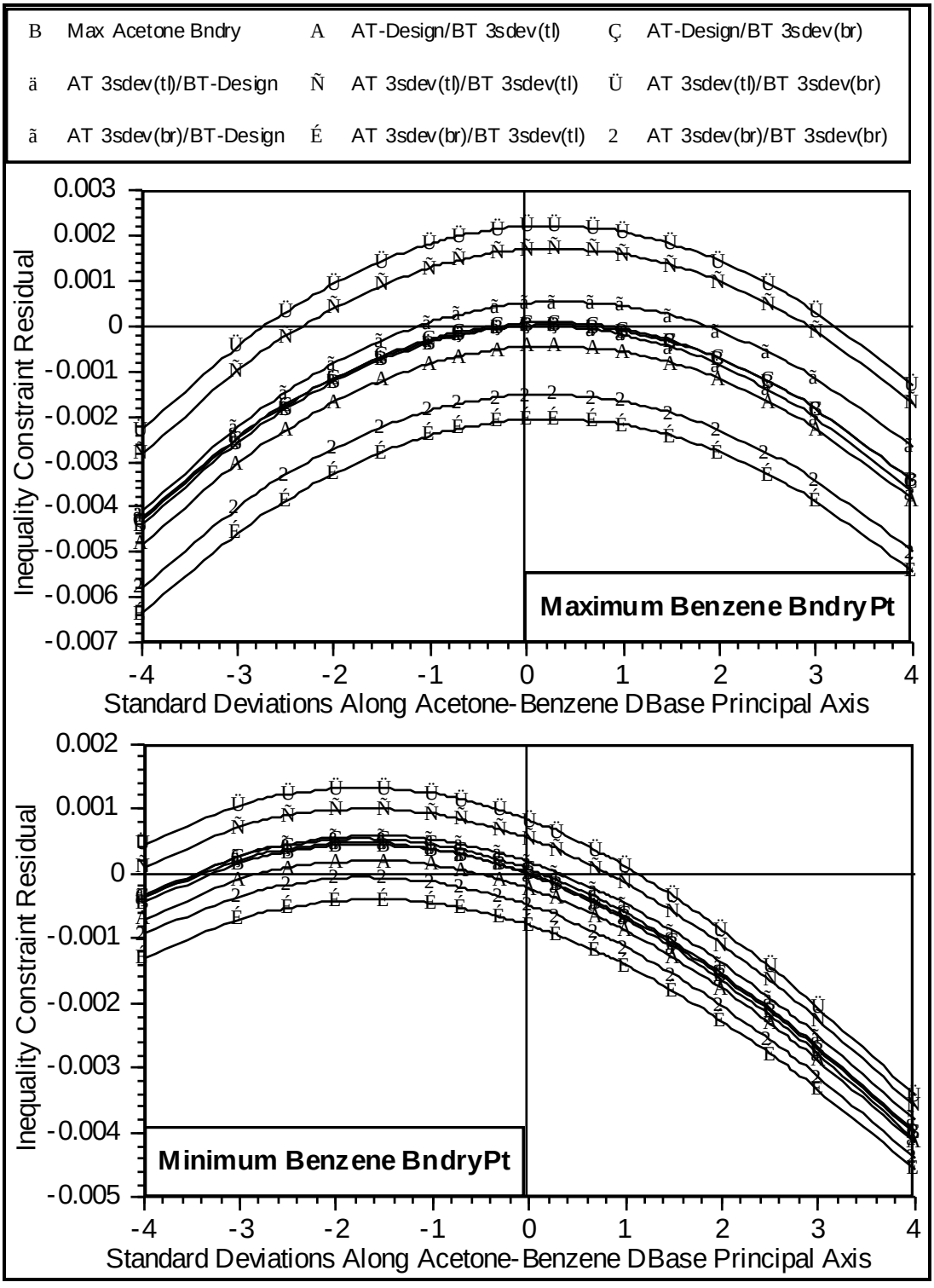

B Max Acetone Bndry
A AT-Design/BT 3sdev(tl)
Ç AT-Design/BT 3sdev(br)
ä AT 3sdev(tl)/BT-Design
Ñ AT 3sdev(tl)/BT 3sdev(tl)
Ü AT 3sdev(tl)/BT 3sdev(br)
ã AT 3sdev(br)/BT-Design
É AT 3sdev(br)/BT 3sdev(tl)
2 AT 3sdev(br)/BT 3sdev(br)
Inequality Constraint Residual
Maximum Benzene Bndry Pt
Standard Deviations Along Acetone-Benzene DBase Principal Axis
Inequality Constraint Residual
Minimum Benzene Bndry Pt
Standard Deviations Along Acetone-Benzene DBase Principal Axis

Figure C-4.d.ii-2 b
Shape of Inequality Constraint as a Function of the Database Uncertainty Parameters Along Their Principal Axes

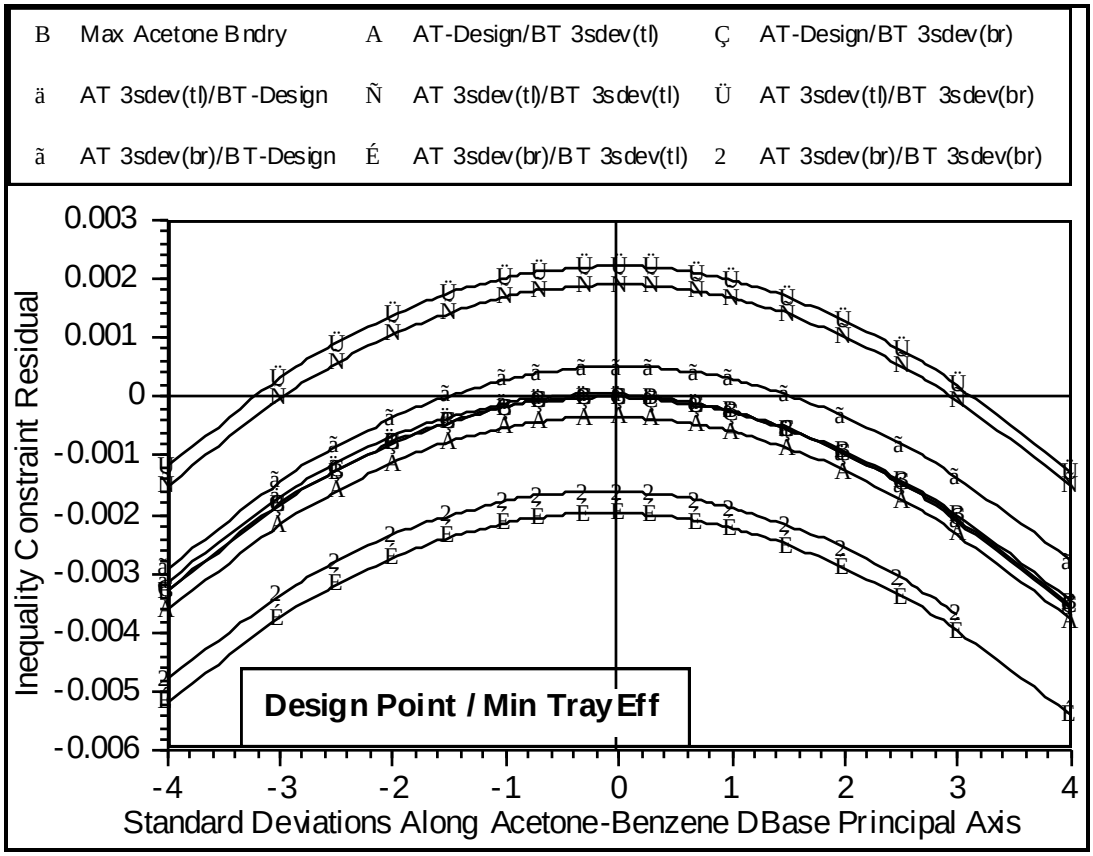


Figure C-4.d.ii-2 c
Shape of Inequality Constraint as a Function of the Database Uncertainty Parameters Along Their Principal Axes

Appendix D. Convergence Plots of Boundary-Approximation Procedure With Distillation Test Problems

This section presents plots to show the convergence characteristics of the boundary-approximation procedure.

D-1. Convergence of the Optimistic – Pessimistic Reliability Estimates

This section presents a plot showing that convergence of the optimistic–pessimistic reliability estimates is log-log. Increased dimension slows convergence rate.

The optimistic and pessimistic boundary-approximations converge asymptotically towards each other in a log-log relationship as a function of the number of boundary points. This is shown in Figure D-1-1 for the acetone-benzene, acetone-water, and 2MB1-IPM test problems which were converged with 10^5 Monte Carlo integration points. Clearly, a strong systematic relationship exists between the convergence rate and the parameter dimension. The differences between test problems is of lesser importance, which indicates that the type of parameter dimension, i.e., statistically significant or insignificant, linear or nonlinear, is not crucial.

For comparison, the convergence rate of a 95% confidence interval for Conventional Monte Carlo (true reliability = 95%) is also plotted. Its convergence rate is slower than the feed+tray.efficiency case, but faster than the feed+tray.efficiency+database case.

While the optimistic boundary-approximation is generally quite accurate, the pessimistic boundary-approximation is overly pessimistic and slow to improve. As dimension increases, the initial pessimistic reliability estimate becomes ridiculously low. For example, for p=5 it is less than 10% for a 90% reliability design. For p=10, it is 0% and had not significantly increased even after 300 boundary points. If the three-component case were plotted in Figure D-1-1, the slope would be practically horizontal.

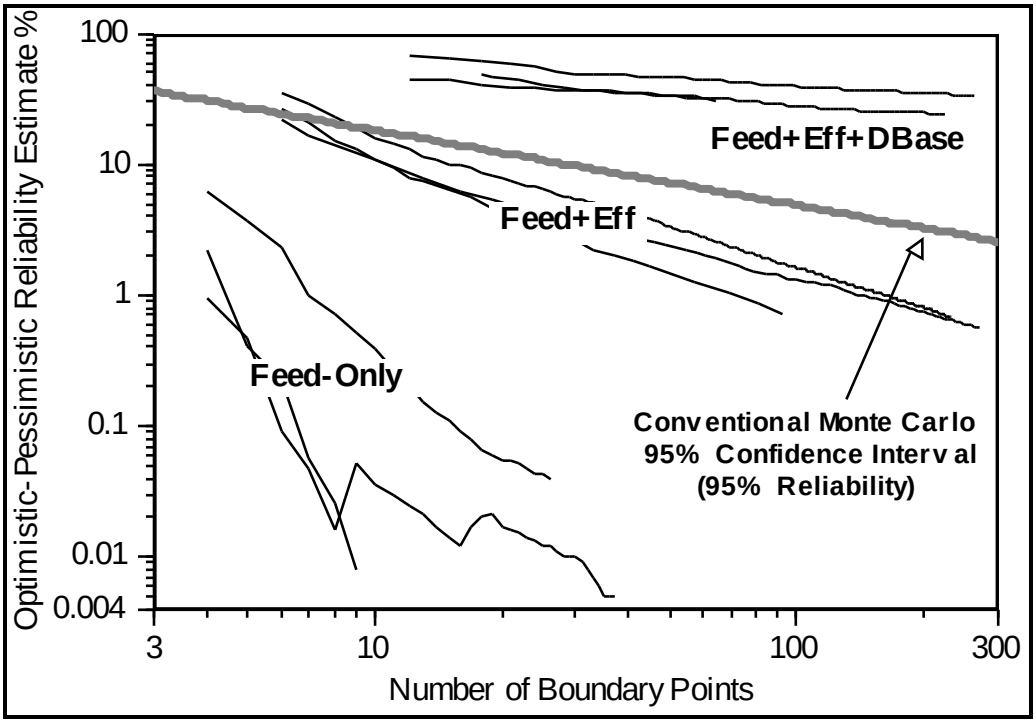


Figure D-1-1
Convergence of the Optimistic–Pessimistic Reliability Estimates vs. the Number of Boundary Points: Two-Component Test Problems

D-2. Convergence of the Best-Estimate Reliability

This section presents plots of the converging best-estimate reliability .

This section plots the convergence of the best-estimate reliability versus different factors in an attempt to establish some clear functional relationship, which might then be used to monitor and accelerate convergence. The factors examined are: (1) number of boundary points, (2) number of total process simulations, (3) number of boundary failure points. The first is of geometrical relevance. The second relates to the process simulation effort. The third is perhaps the most geometrically relevant because it ignores the nonfailure boundary points which do nothing to directly improve the best-estimate reliability.

Both log-log and log-linear plots are presented. Convergence is at least log-log, because the best-estimate reliability is bounded by the optimistic and pessimistic reliability estimates. In these plots, whichever slope is most constant indicates the better relationship. 10^5 Monte Carlo integration points were used to provide smoother slopes and greater resolution.

D-2.a. Acetone-Benzene

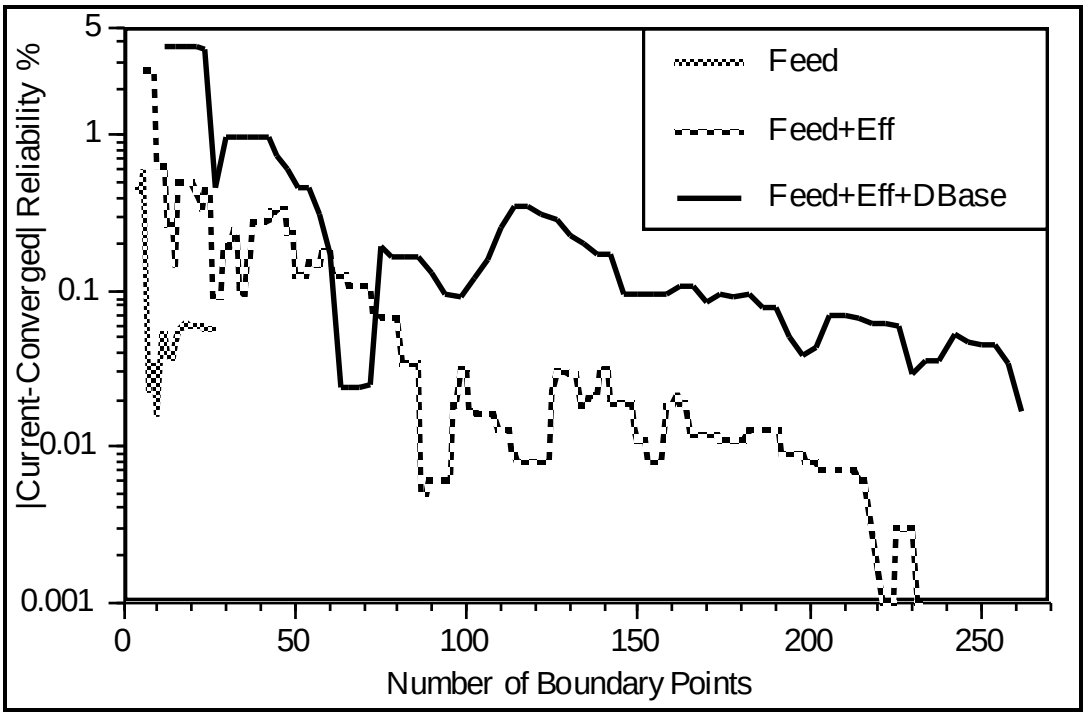


Figure D-2.a.1. Acetone-Benzene. Log-Linear Relationship of Convergence vs. Number of Boundary Points

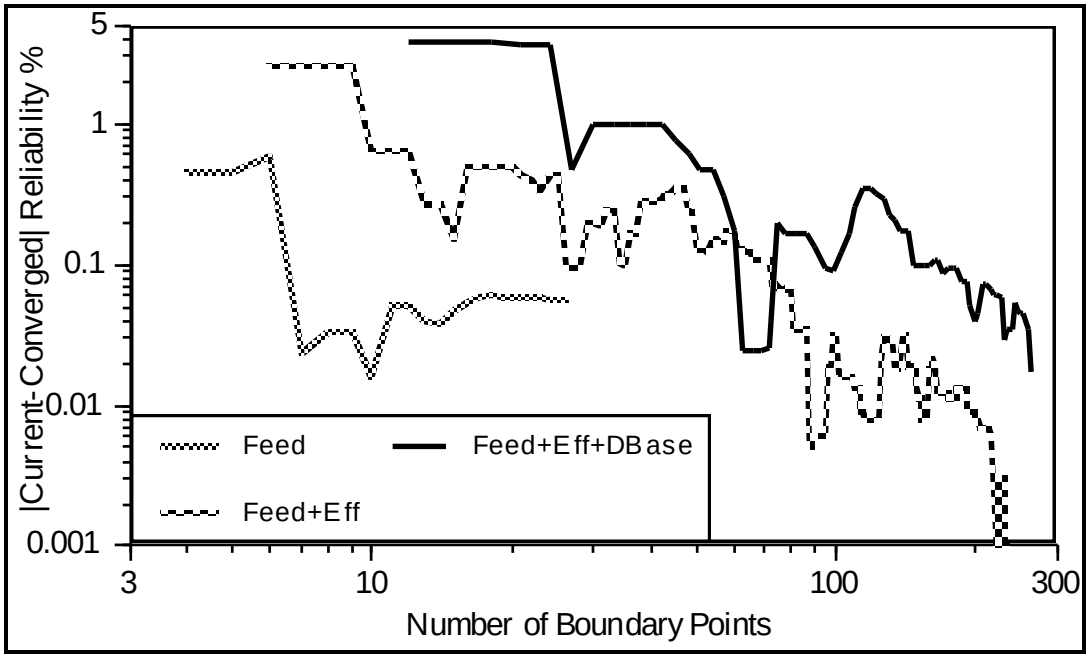


Figure D-2.a.2. Acetone-Benzene. Log-Log Relationship of Convergence vs. Number of Boundary Points

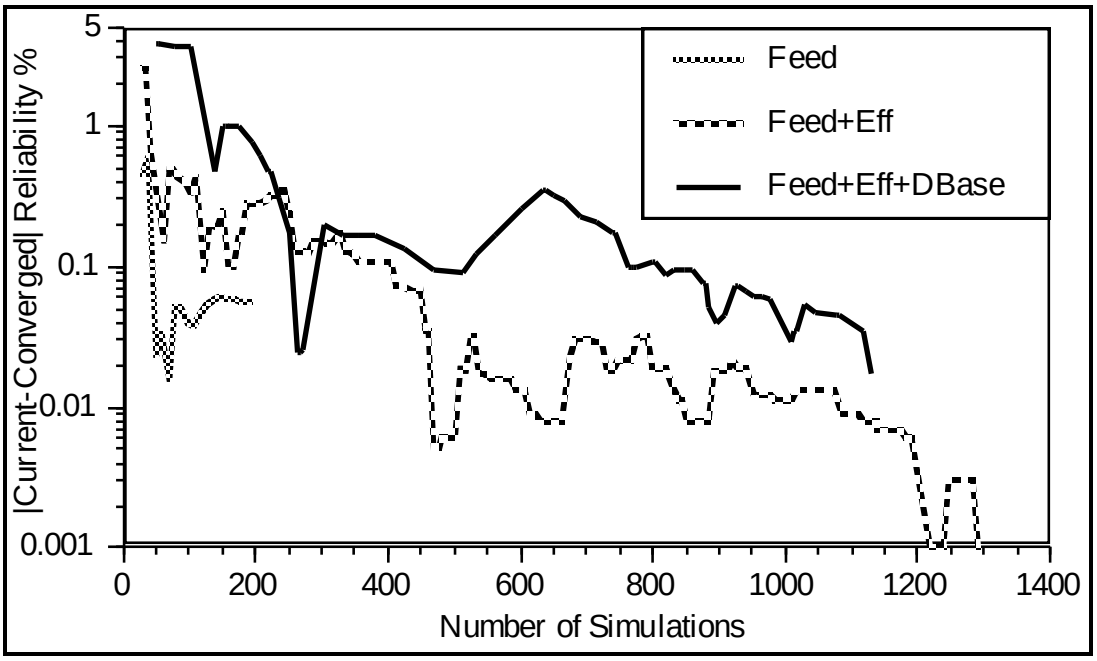


Figure D-2.a.3. Acetone-Benzene. Log-Linear Relationship of Convergence vs. Number of Simulations

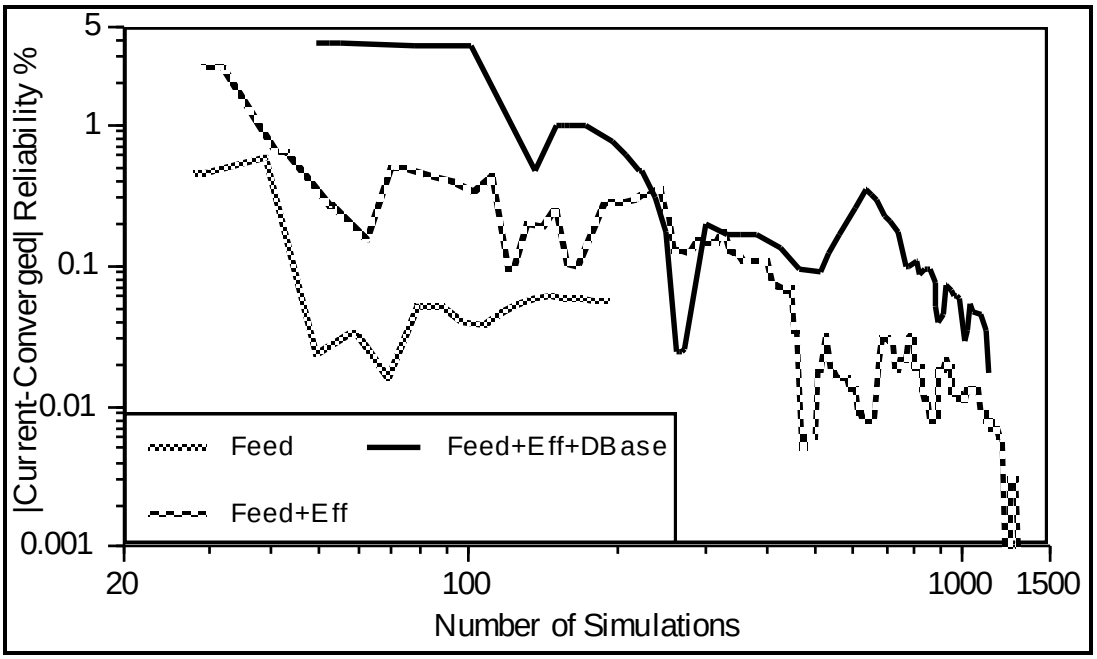


Figure D-2.a.4. Acetone-Benzene. Log-Log Relationship of Convergence vs. Number of Simulations

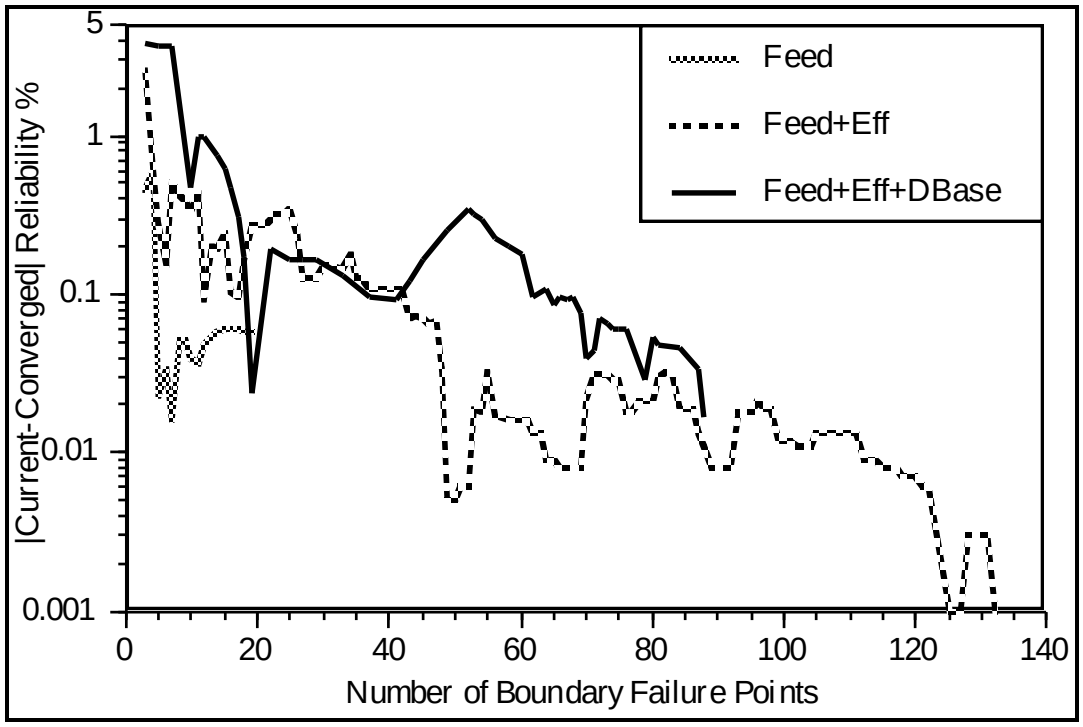


Figure D-2.a.5. Acetone-Benzene. Log-Linear Relationship of Convergence vs. Number of Boundary Failure Points

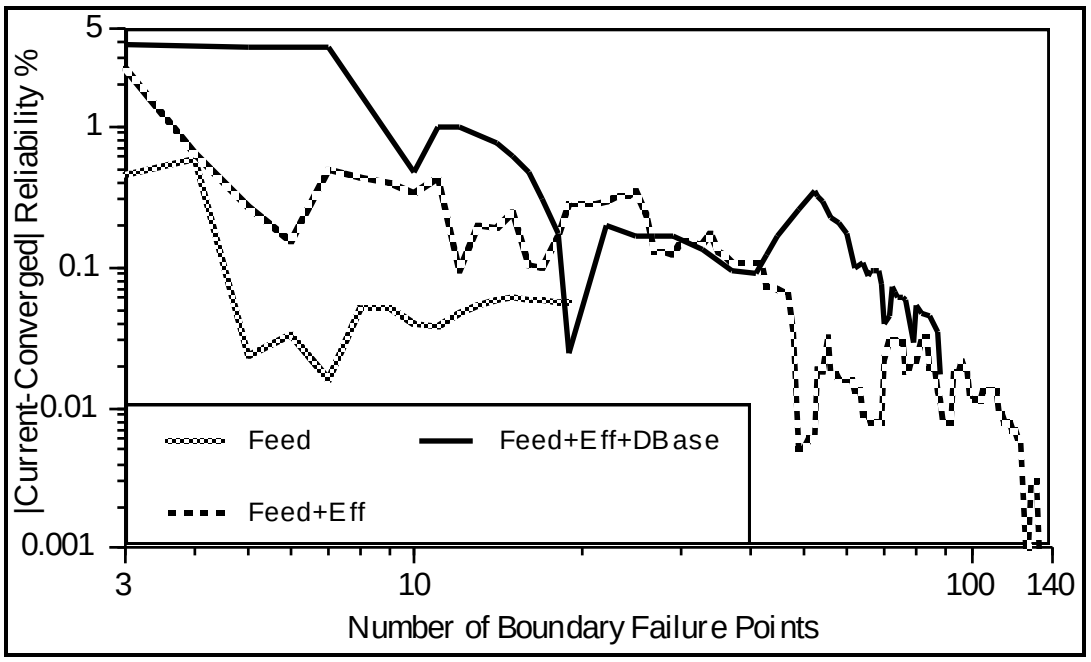


Figure D-2.a.6. Acetone-Benzene. Log-Log Relationship of Convergence vs. Number of Boundary Failure Points

D-2.b. Acetone-Water

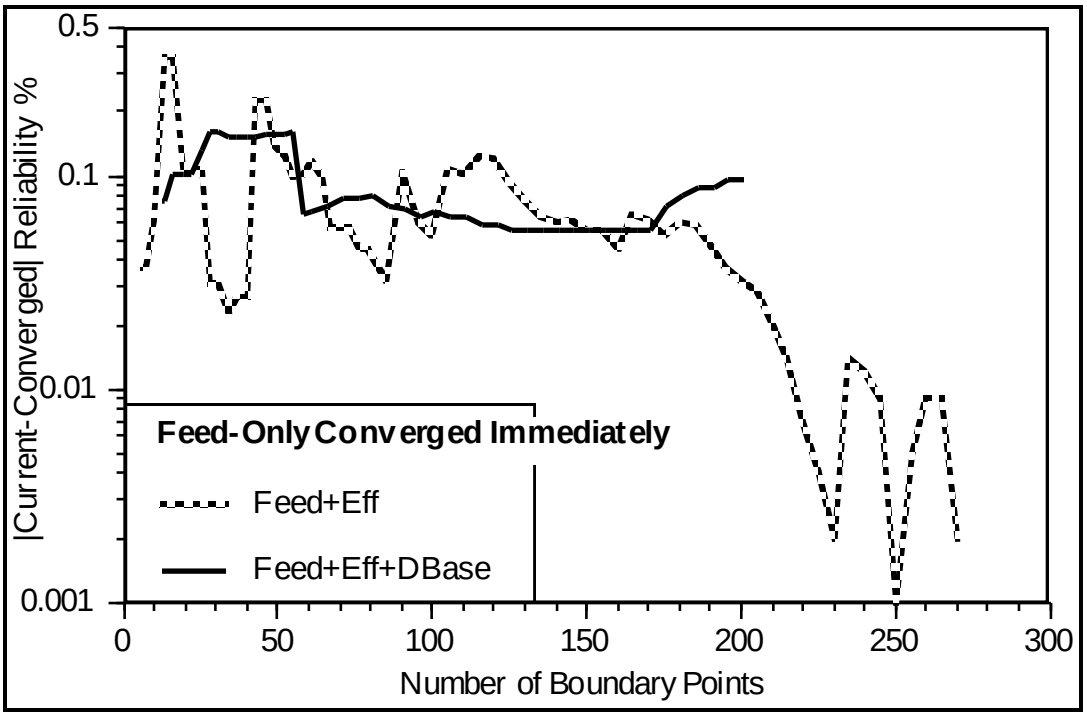


Figure D-2.b.1. Acetone-Water. Log-Linear Relationship of Convergence vs. Number of Boundary Points

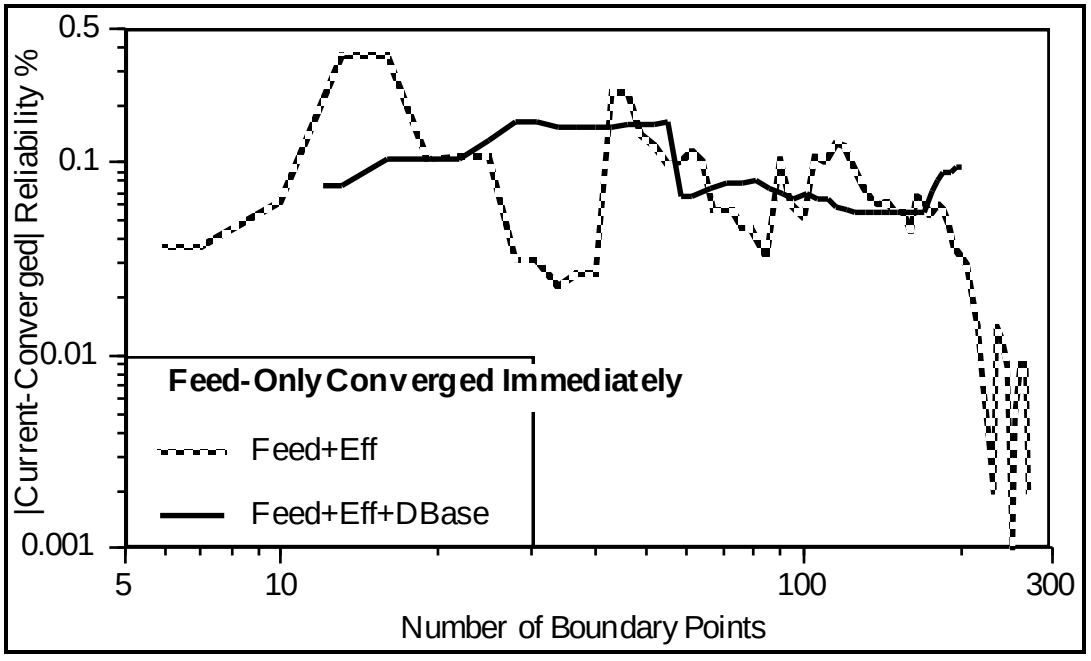


Figure D-2.b.2. Acetone-Water. Log-Log Relationship of Convergence vs. Number of Boundary Points

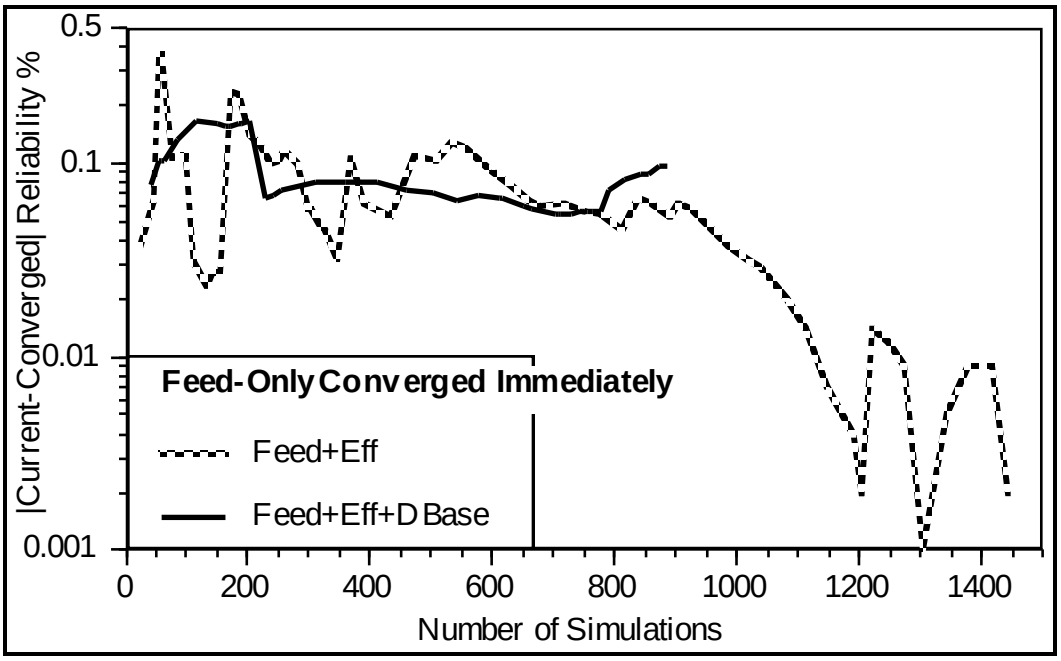


Figure D-2.b.3. Acetone-Water. Log-Linear Relationship of Convergence vs. Number of Simulations

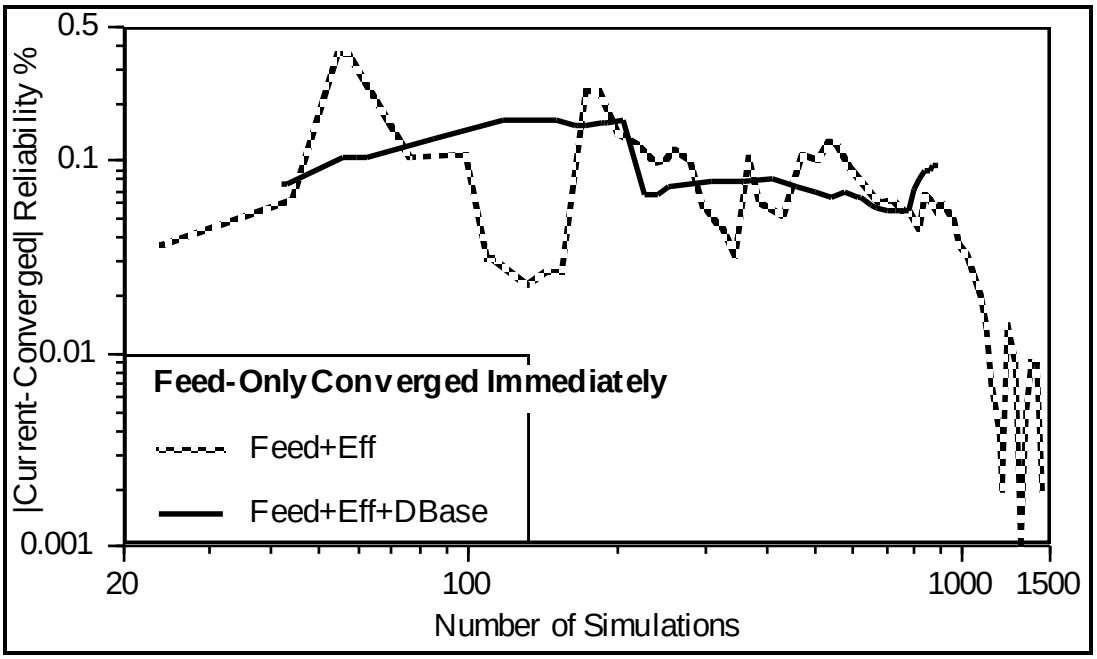


Figure D-2.b.4. Acetone-Water. Log-Log Relationship of Convergence vs. Number of Simulations

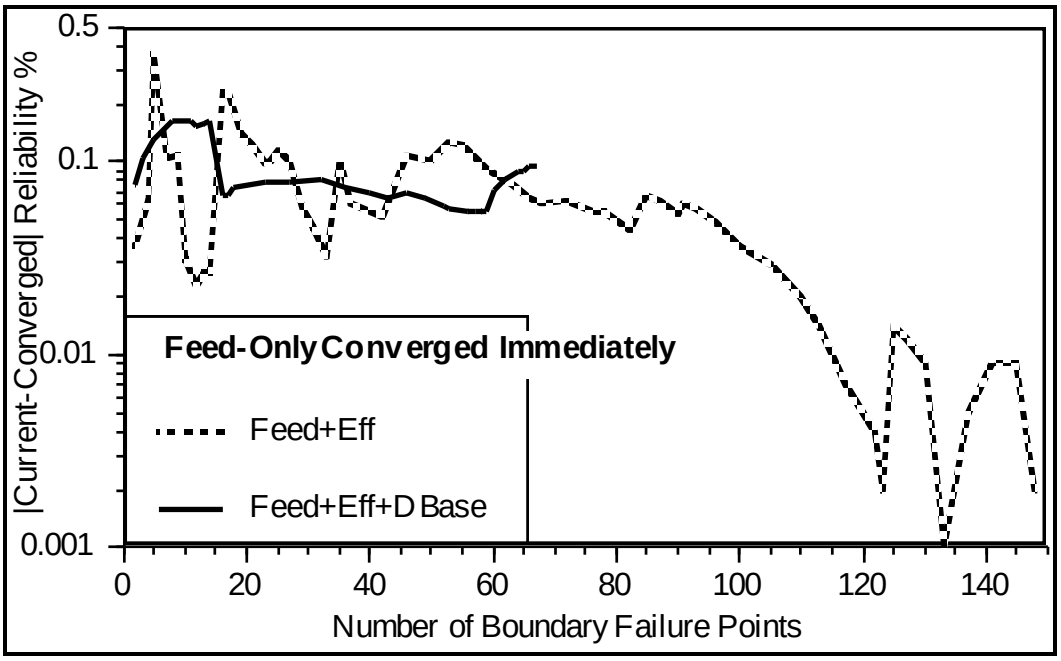


Figure D-2.b.5. Acetone-Water. Log-Linear Relationship of Convergence vs. Number of Boundary Failure Points

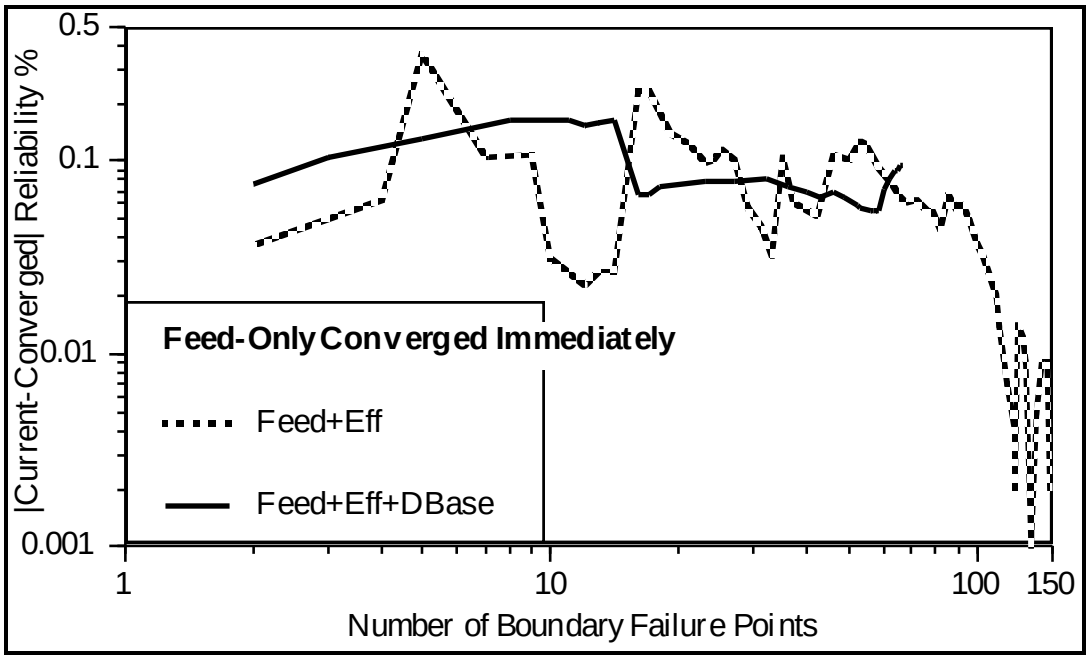


Figure D-2.b.6. Acetone-Water. Log-Log Relationship of Convergence vs. Number of Boundary Failure Points

D-2.c. 2MB1-IPM

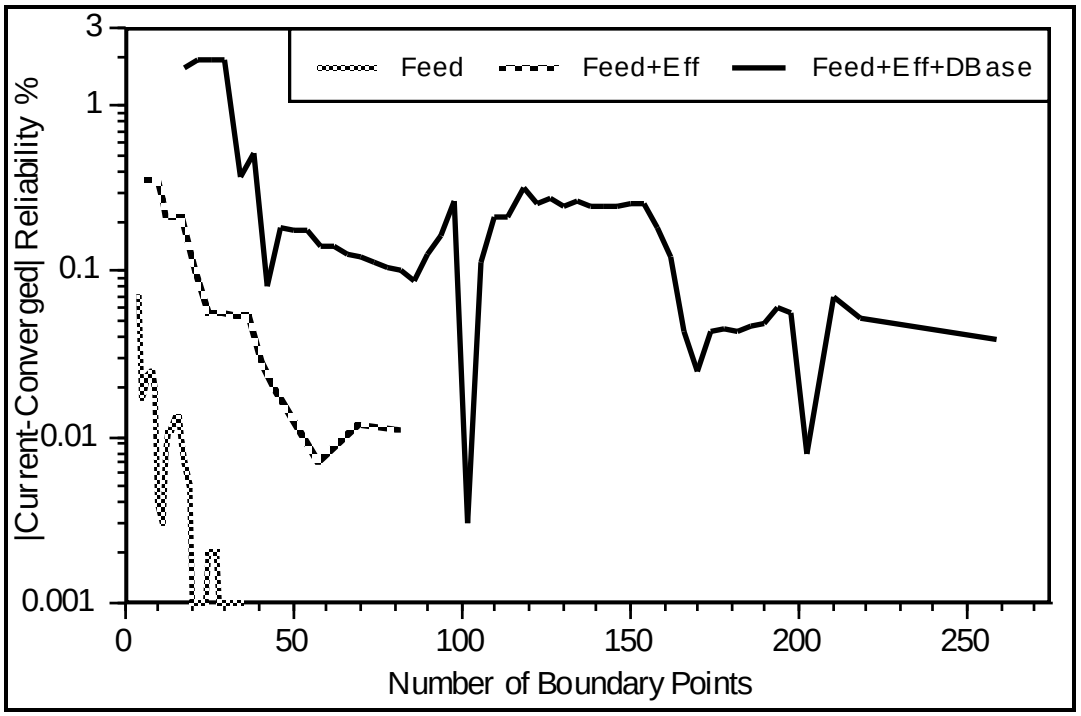


Figure D-2.c.1. 2MB1-IPM. Log-Linear Relationship of Convergence vs. Number of Boundary Points

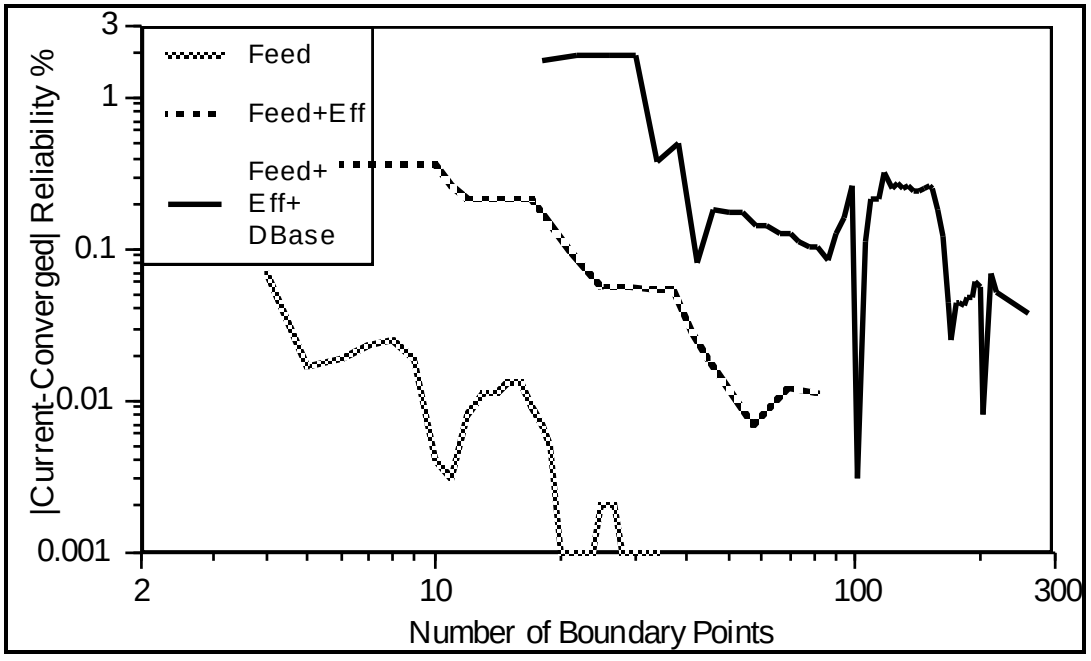


Figure D-2.c.2. 2MB1-IPM. Log-Log Relationship of Convergence vs. Number of Boundary Points

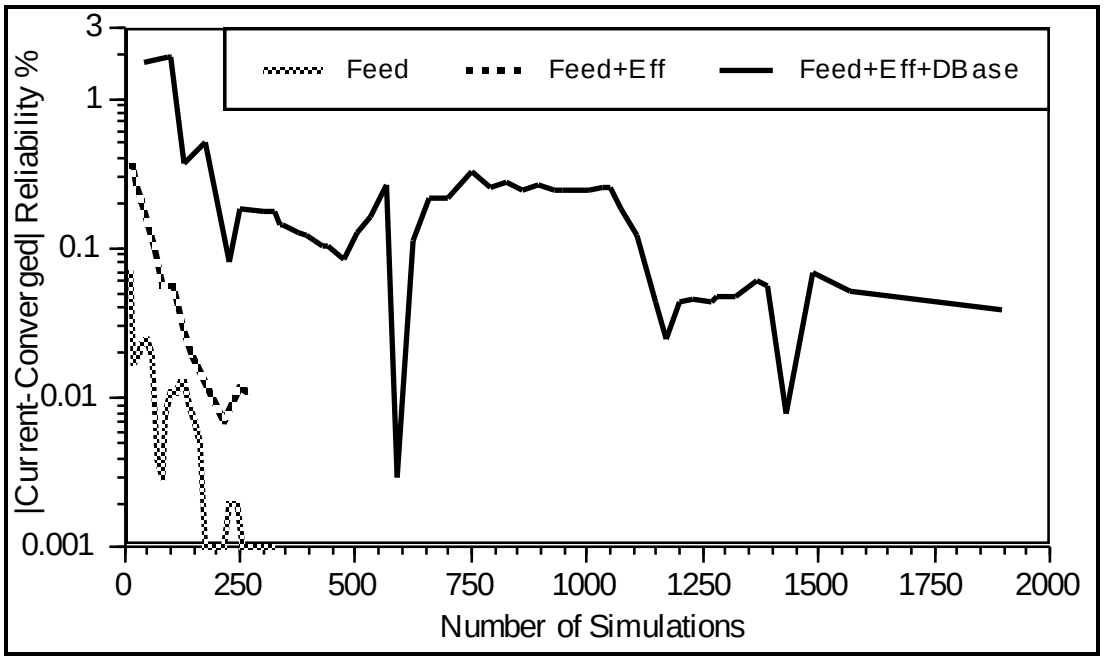


Figure D-2.c.3. 2MB1-IPM. Log-Linear Relationship of Convergence vs. Number of Simulations

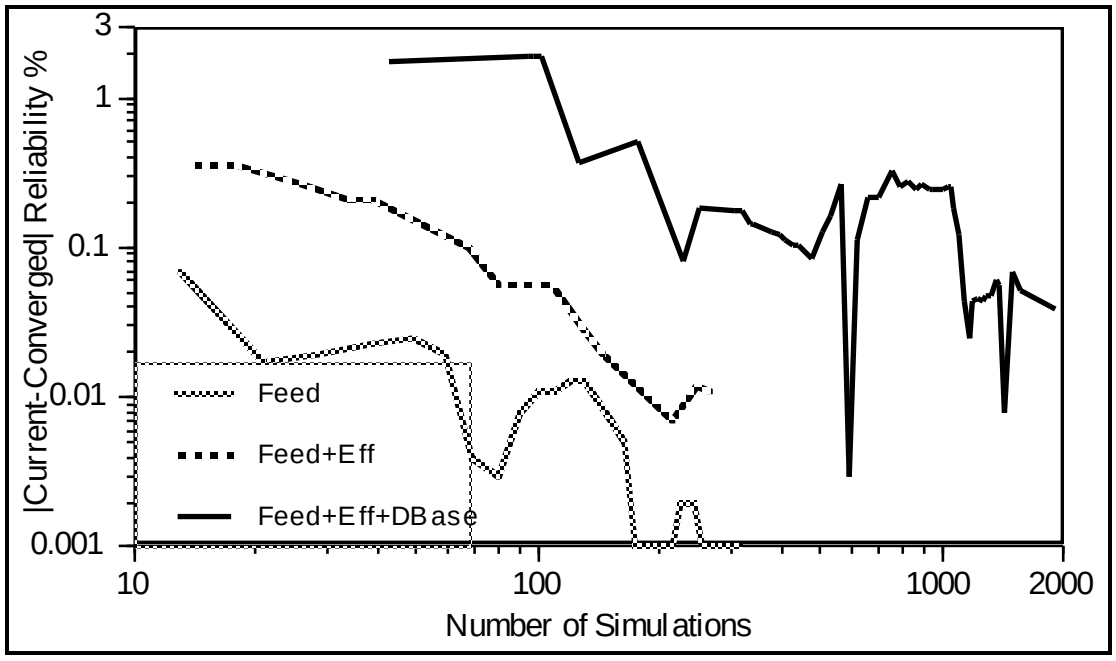


Figure D-2.c.4. 2MB1-IPM. Log-Log Relationship of Convergence vs. Number of Simulations

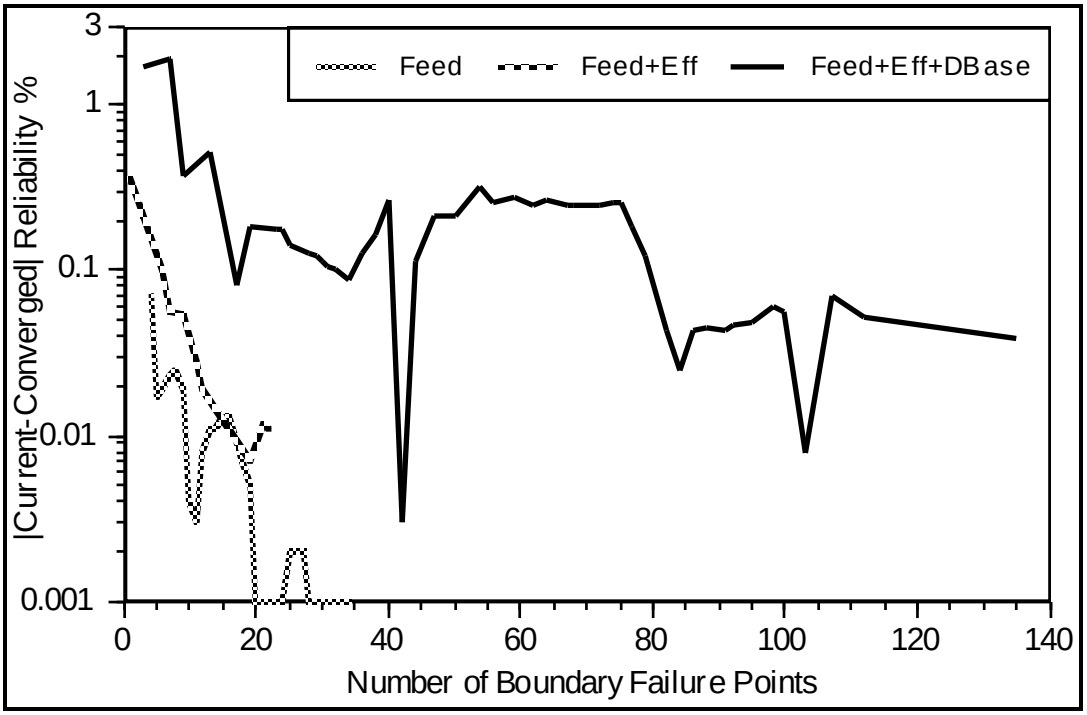


Figure D-2.c.5. 2MB1-IPM. Log-Linear Relationship of Convergence vs. Number of Boundary Failure Points

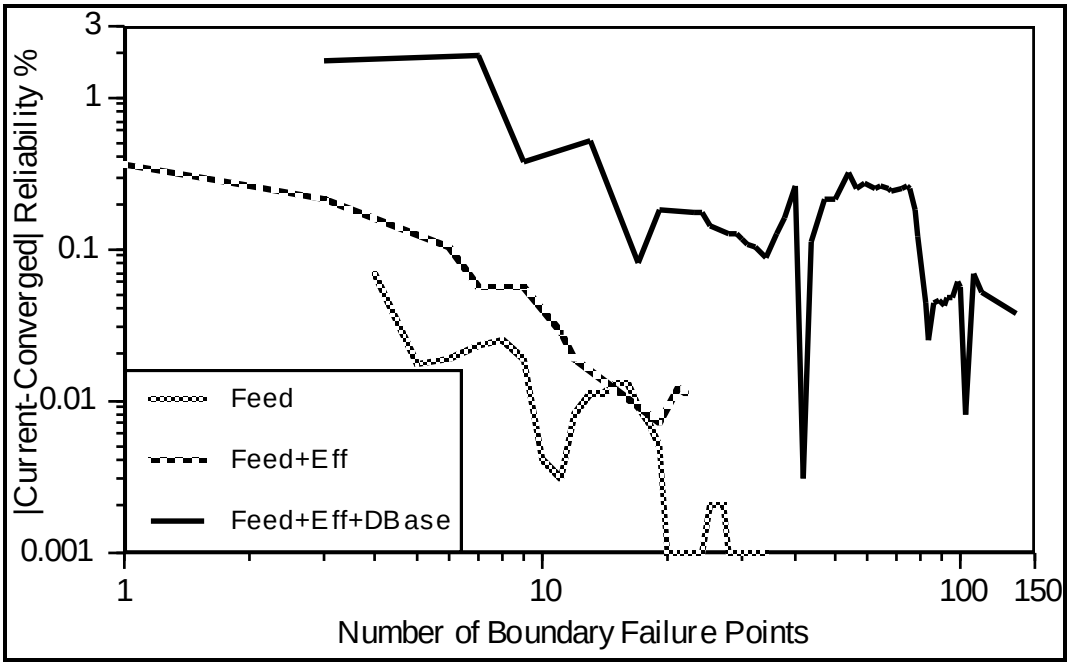


Figure D-2.c.6. 2MB1-IPM. Log-Log Relationship of Convergence vs. Number of Boundary Failure Points

D-2.d. Acetone-Benzene-Toluene

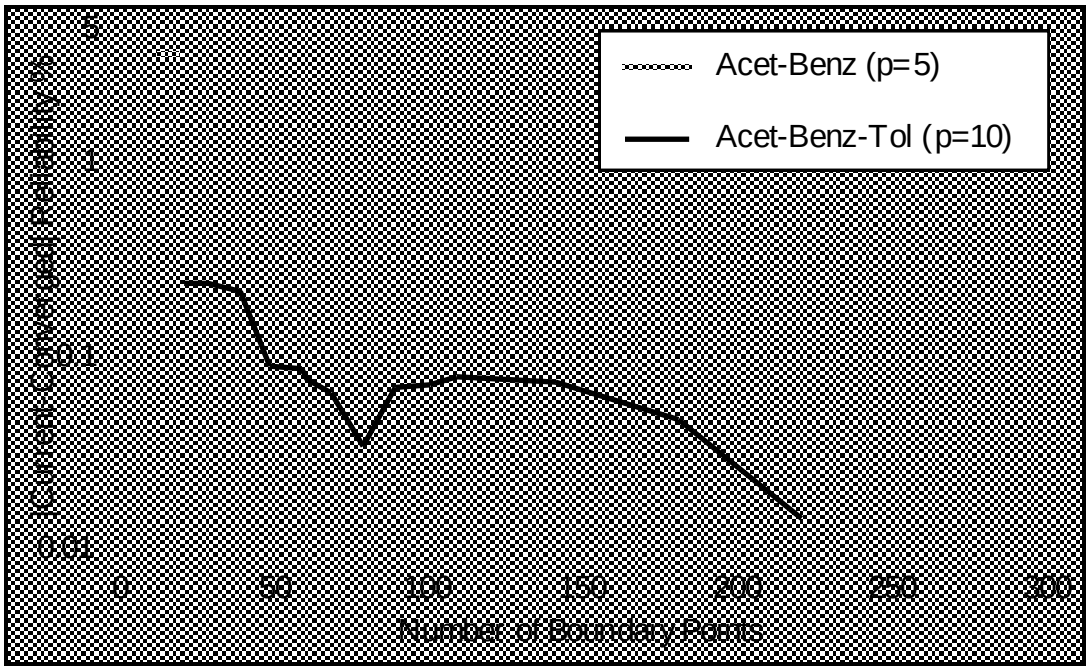


Figure D-2.d.1. Acetone-Benzene-Toluene. Log-Linear Relationship of Convergence vs. Number of Boundary Points

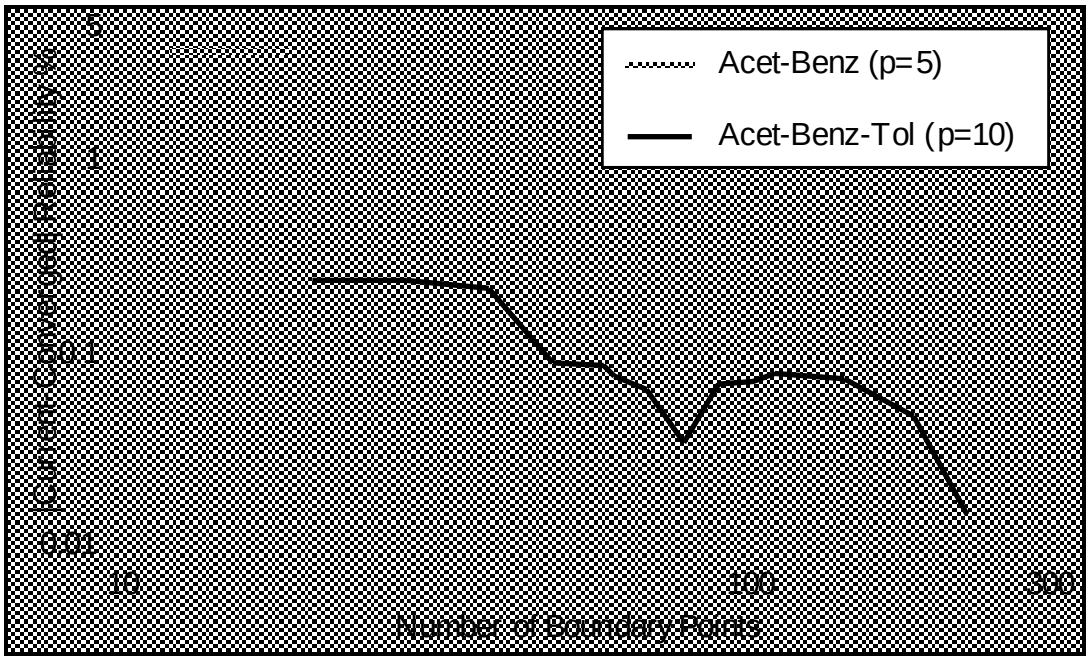


Figure D-2.d.2. Acetone-Benzene-Toluene. Log-Log Relationship of Convergence vs. Number of Boundary Points

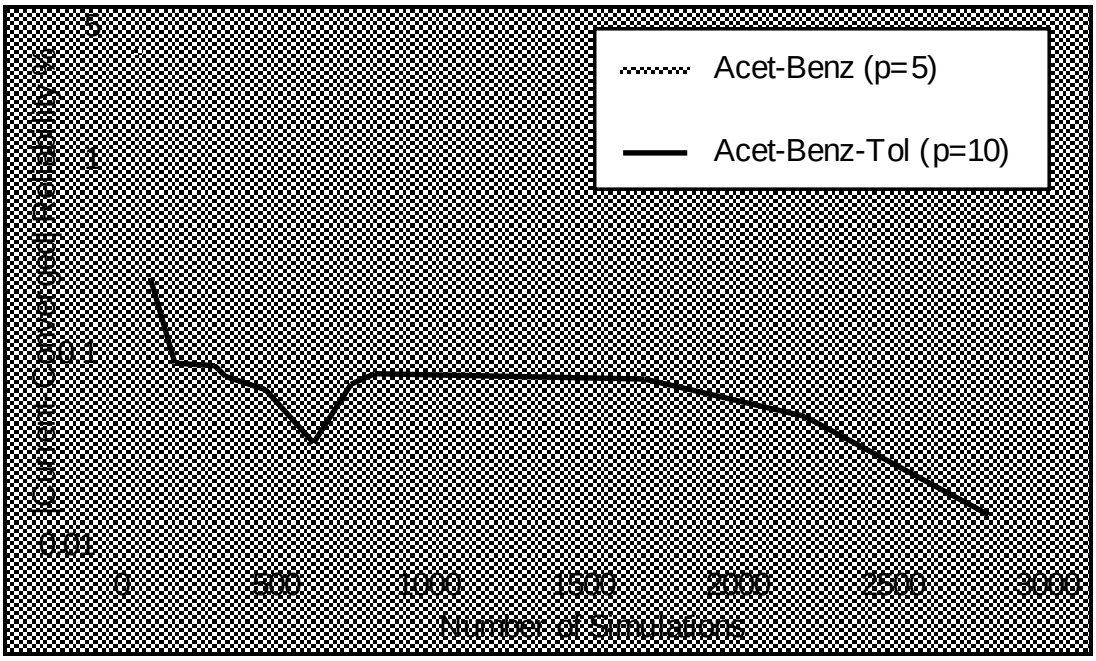


Figure D-2.d.3. Acetone-Benzene-Toluene. Log-Linear Relationship of Convergence vs. Number of Simulations

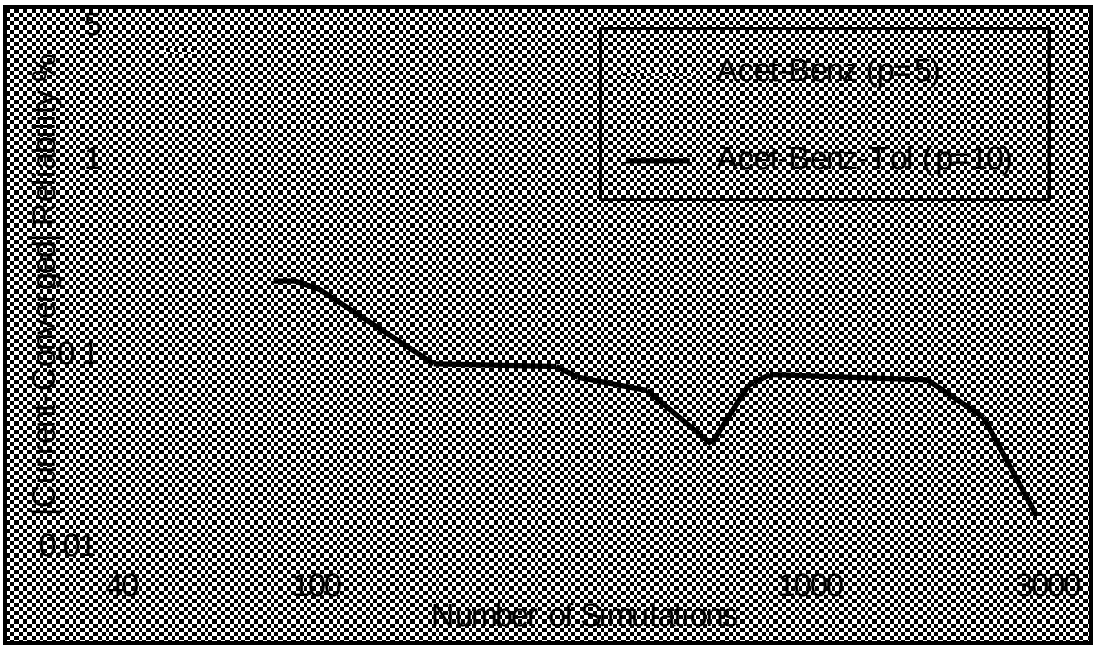


Figure D-2.d.4. Acetone-Benzene-Toluene. Log-Log Relationship of Convergence vs. Number of Simulations

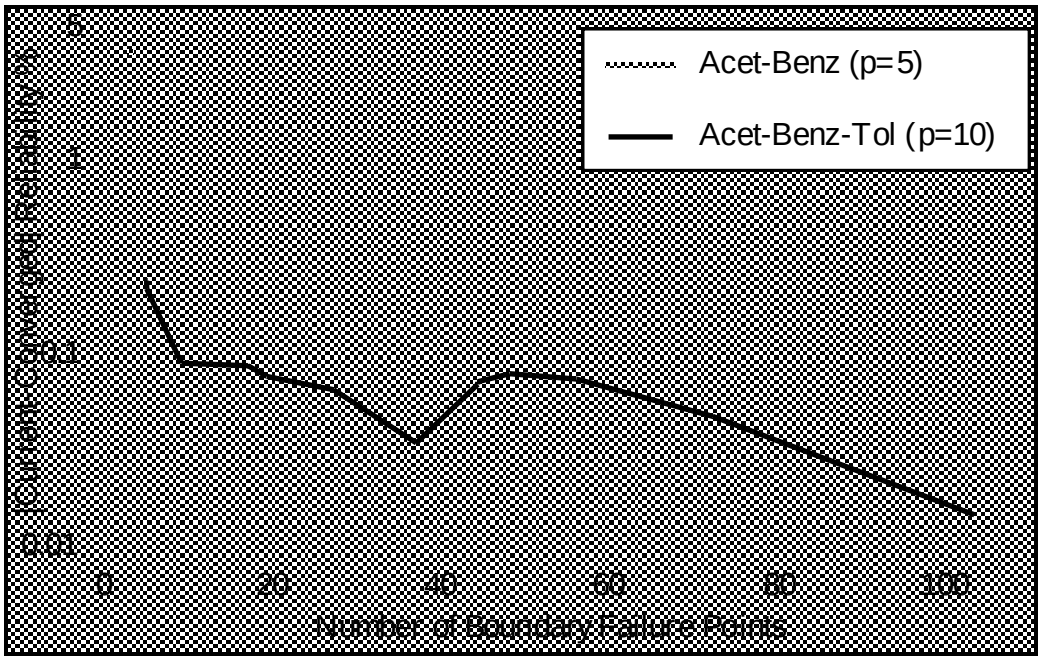


Figure D-2.d.5. Acetone-Benzene-Toluene. Log-Linear Relationship of Convergence vs. Number of Boundary Failure Points

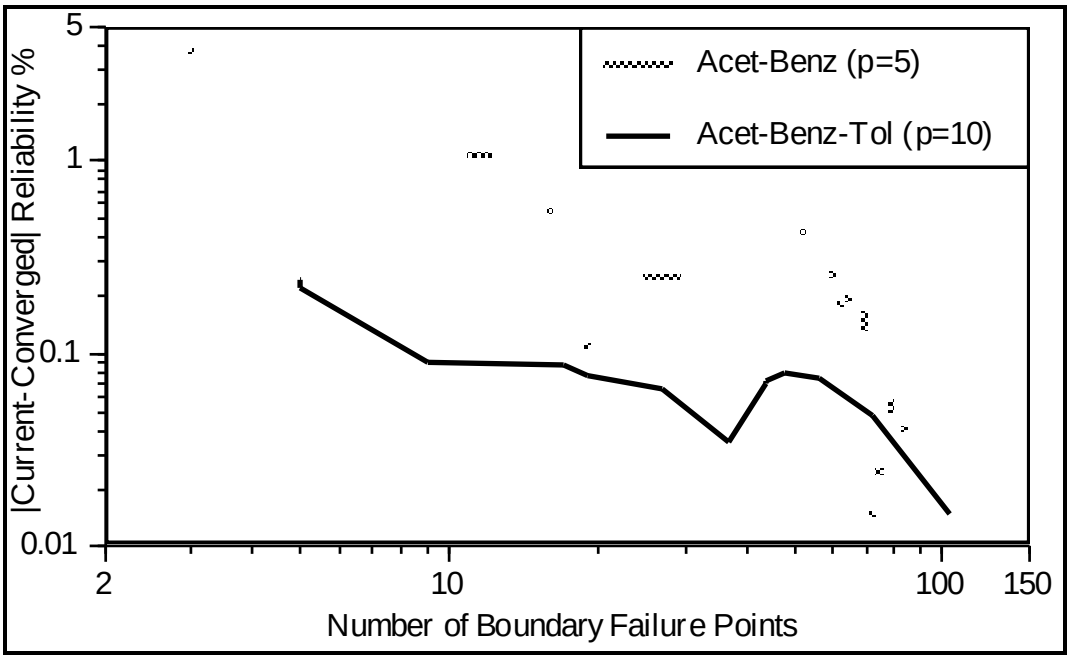


Figure D-2.d.6. Acetone-Benzene-Toluene. Log-Log Relationship of Convergence vs. Number of Boundary Failure Points

D-2.e. Observations

Although the evidence is inconclusive, generally the log-linear plots show slight upward curvature, while the log-log plots show straight or downward curvature. This indicates convergence is closer to log-log than log-linear. However, there are exceptions to this trend. When initial accuracy is high, e.g., acetone-water, convergence is less systematic.

None of the three factors (boundary points, process simulations, failure boundary points) is clearly superior for tracking convergence.

The information is insufficient to clearly determine the effect of parameter dimension on convergence. Slopes are fairly constant with dimension, indicating that convergence rate does not change with dimension. However, the greater initial inaccuracy and time required for initialization slow convergence by shifting the convergence slope to the ‘right’ (longer times).

Appendix E. Comparison Plots of the Boundary-Approximation Procedure versus Conventional Monte Carlo

This section replots the figures presented in sections V.C.2.b. and V.C.2.c, except with each test problem shown separately for clarity.

E-1. Execution Time-Ratio (95% Confidence Interval of Conventional Monte Carlo / Boundary-Approximation) as a Function of Accuracy in the Reliability Estimate

This section repeats Figure V.C.2.b-1, except each test problem is separated into individual plots. These plots are intended to show that the boundary-approximation procedure is orders of magnitude faster than Conventional Monte Carlo. They are also intended to highlight the effects of parameter dimension, if any.

Note that the 95% confidence interval for Conventional Monte Carlo is not symmetrical, so accuracy was taken as (Minimum%—Minimum%) / 2 instead. Also, because convergence was often noisy, only the minimum and maximum execution time-ratios are plotted at each 0.1% accuracy interval (the maximum possible resolution).

It was reported that the boundary-approximation procedure is 20-7000 times faster than Conventional Monte Carlo. These are reasonable values with the higher ratios (two to three orders of magnitude) being more indicative because they occur at convergence. The ratio for the acetone-benzene test problem (Figure E-2-1) was actually 4-10 at initialization because of its high error of 2.5% to 5.0%. However, the error was quickly corrected to less than 1.0% error and the ratio increased to 30. The procedure then converged with a time-ratio of 7000.

In all cases, the log-log plots show a linear slope. This is because the procedure converges so fast in relation to Conventional Monte Carlo that the execution time ratio is basically a function only of reliability accuracy. Any effect of parameter dimension will appear as raised or lowered points.

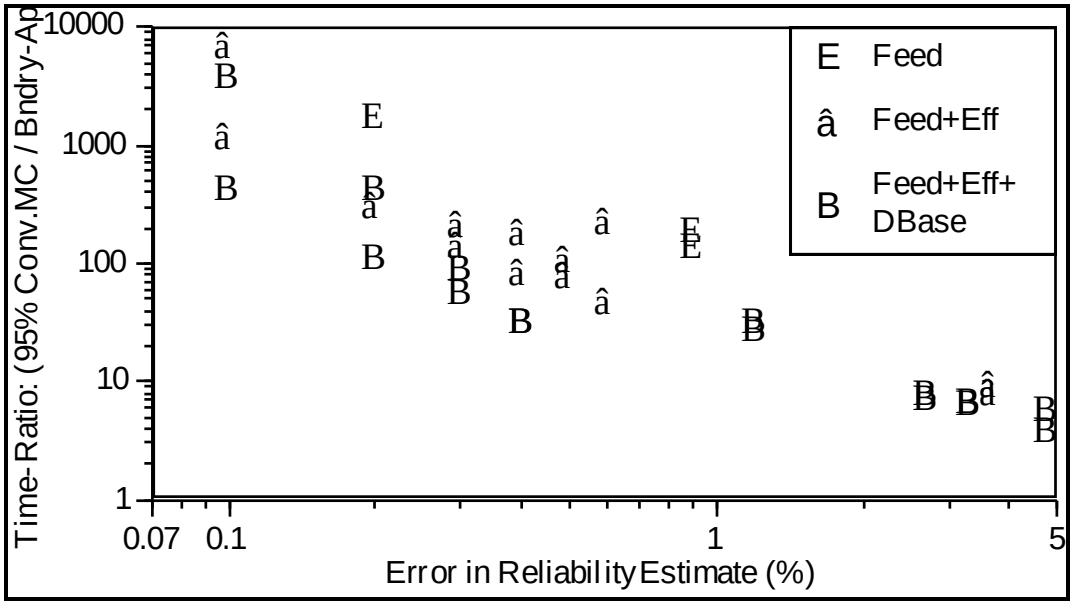


Figure E.2-1
Time Required for 95% Confidence Interval of Conventional Monte Carlo Divided by Time Required for Boundary-Approximation Procedure as a Function of Reliability Estimate Accuracy: Acetone-Benzene

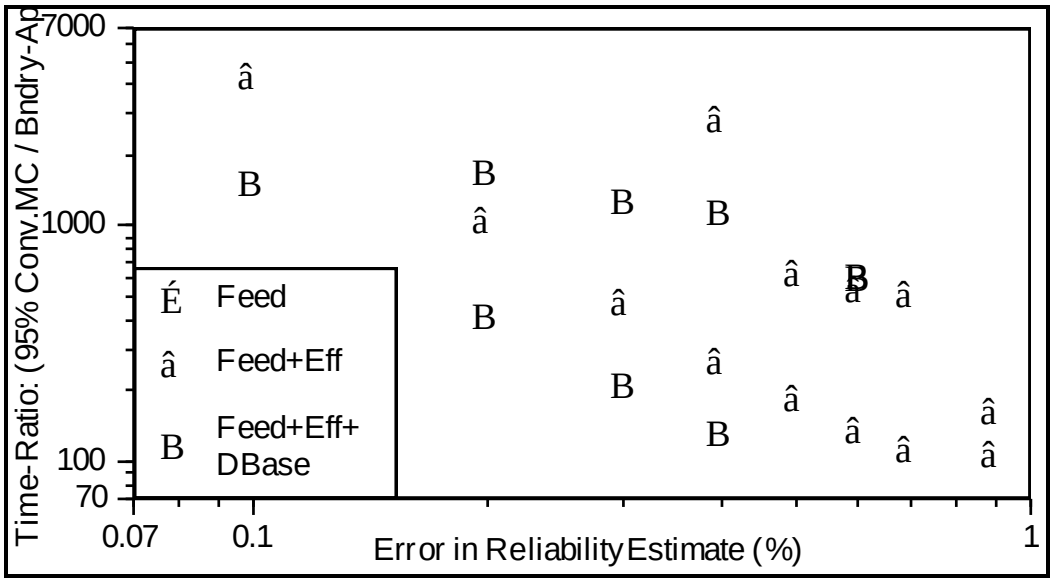


Figure E-1-2
Time Required for 95% Confidence Interval of Conventional Monte Carlo Divided by Time Required for Boundary-Approximation Procedure as a Function of Reliability Estimate Accuracy: Acetone-Water

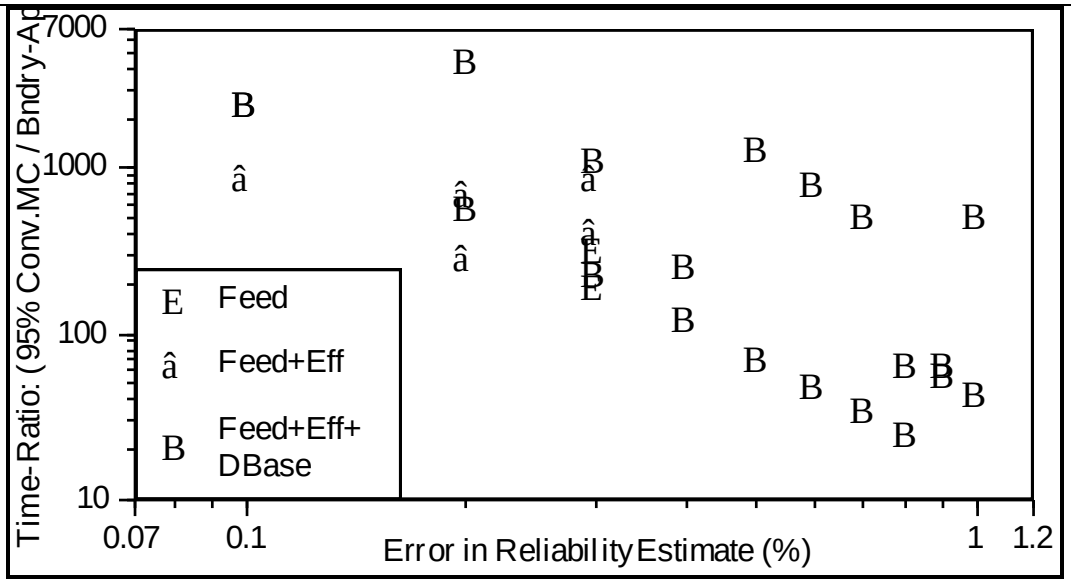


Figure E-1-3
Time Required for 95% Confidence Interval of Conventional Monte Carlo Divided by Time Required for Boundary-Approximation Procedure as a Function of Reliability Estimate Accuracy: 2MB1-IPM

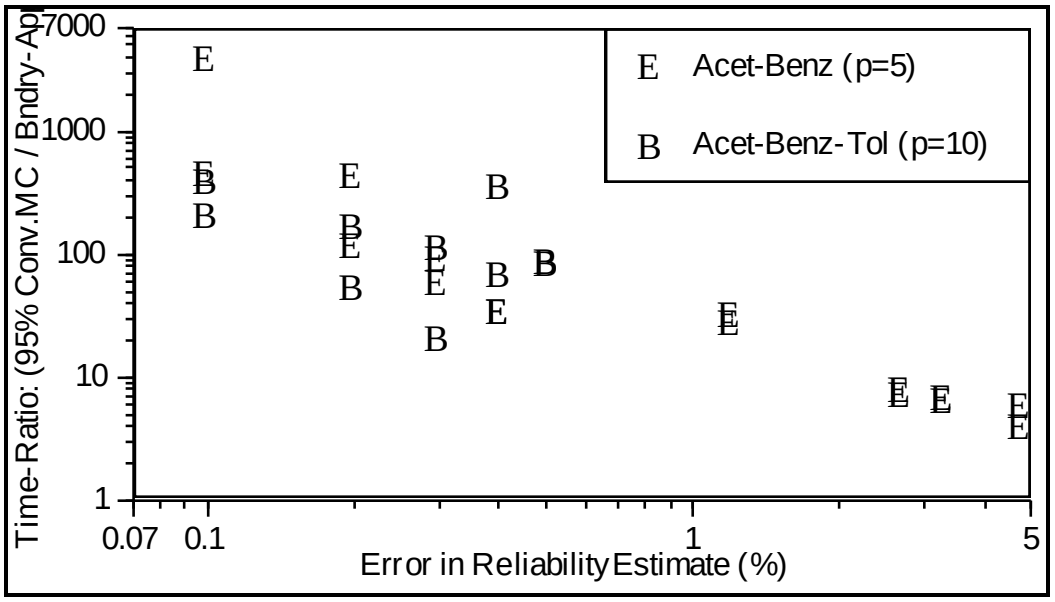


Figure E-1-4
Time Required for 95% Confidence Interval of Conventional Monte Carlo Divided by Time Required for Boundary-Approximation Procedure as a Function of Reliability Estimate Accuracy: Acetone-Benzene + Toluene

E-2. Statistical Odds that the Boundary-Approximation Procedure is More Accurate than Conventional Monte Carlo

This section lists the individual test problem plots that were combined into Figure V.C.2.c-1. That figure showed the statistical odds that the boundary-approximation procedure is more likely to be accurate than a single Conventional Monte Carlo run as a function of execution time. This section also shows the statistical odds as a function of reliability accuracy.

Once again, the boundary-approximation procedure converges so much faster than Conventional Monte Carlo that the log-log plots have basically constant, though noisy, slopes. The positive slope when the x-axis is the execution time shows that the boundary-approximation procedure becomes more and more superior as it converges. The one exception is the 2MB1-IPM feed+tray.efficiency+database test problem which has a final error of 0.6% that shows up as a sudden drop at the end of the curve, see Figure E-2-2. Note that parameter dimension is not significant except that the procedure requires more computational effort to get started as dimension increases.

When the x-axis is the reliability accuracy, the slopes are also constant, though noisy, and confirm the obvious observation that the boundary-approximation is more and more superior, the more accurate it gets. These plots also confirm that parameter dimension has no significance.

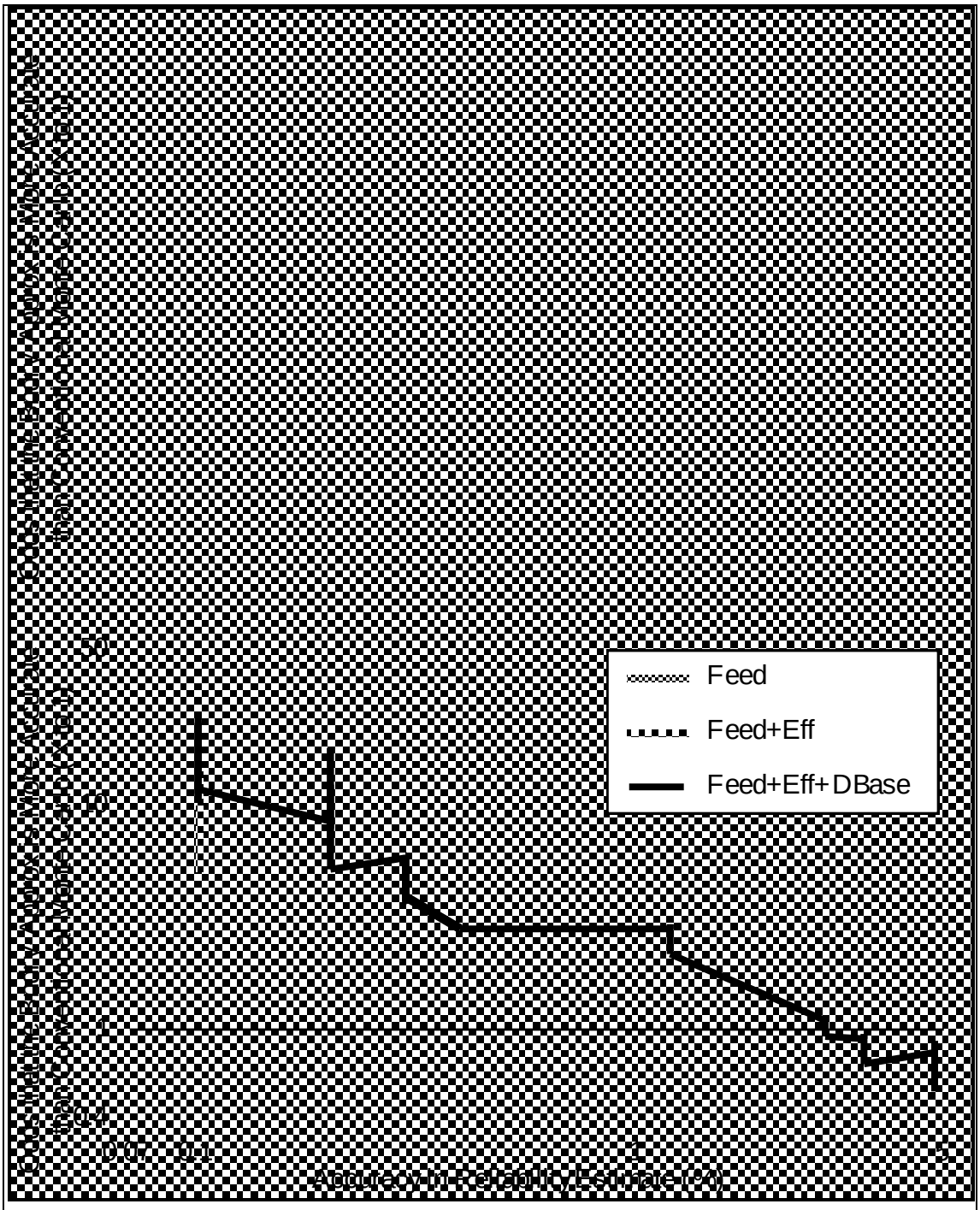


Figure E-2-1
Likelihood that the Boundary-Approximation Procedure is More Accurate than the Conventional Monte Carlo Procedure as a Function of Computational Effort and Accuracy: Acetone-Benzene

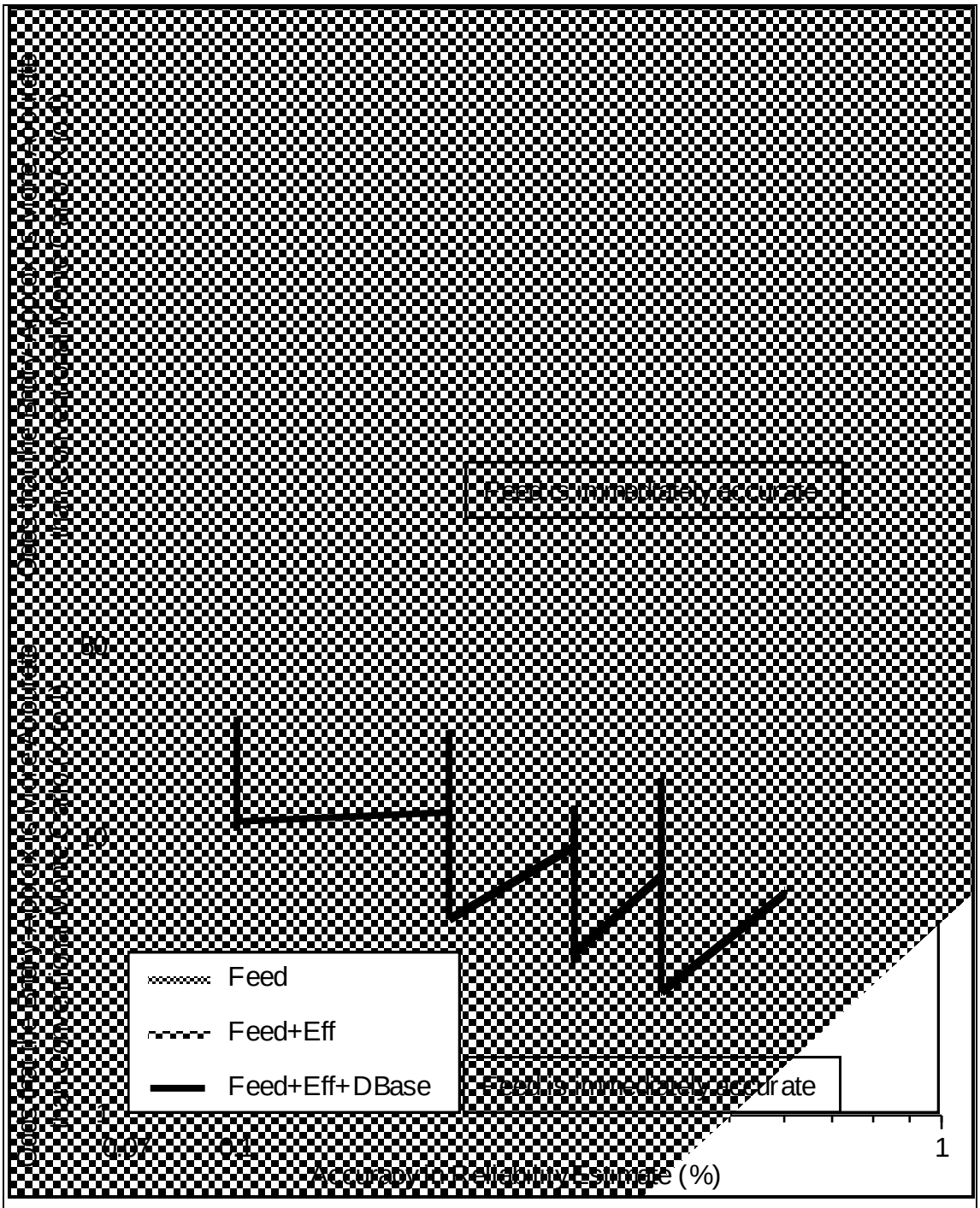


Figure E-2-2
Likelihood that the Boundary-Approximation Procedure is More Accurate than the Conventional Monte Carlo Procedure as a Function of Computational Effort and Accuracy: Acetone-Water

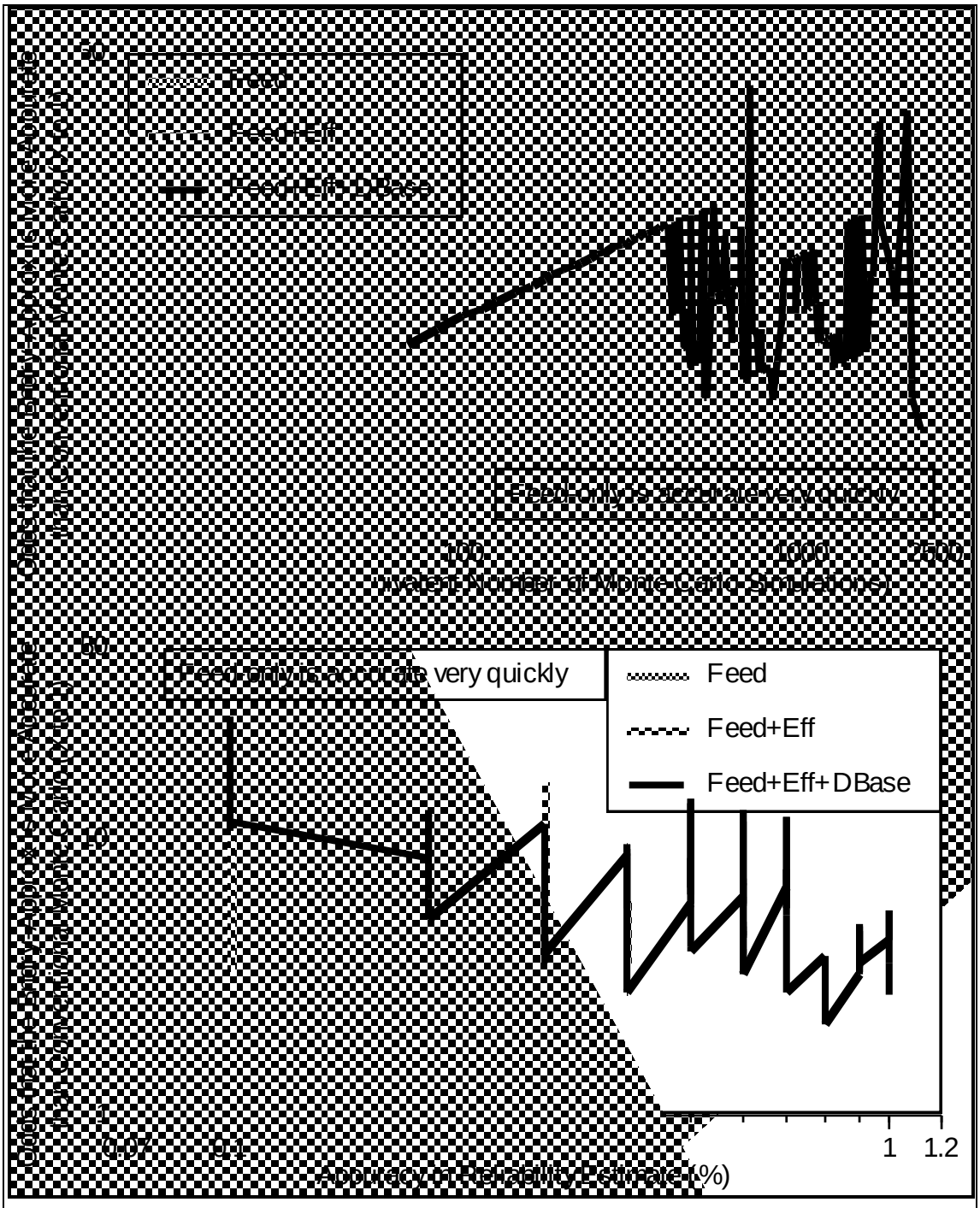


Figure E-2-3
Likelihood that the Boundary-Approximation Procedure is More Accurate than the Conventional Monte Carlo Procedure as a Function of Computational Effort and Accuracy: 2MB1-IPM

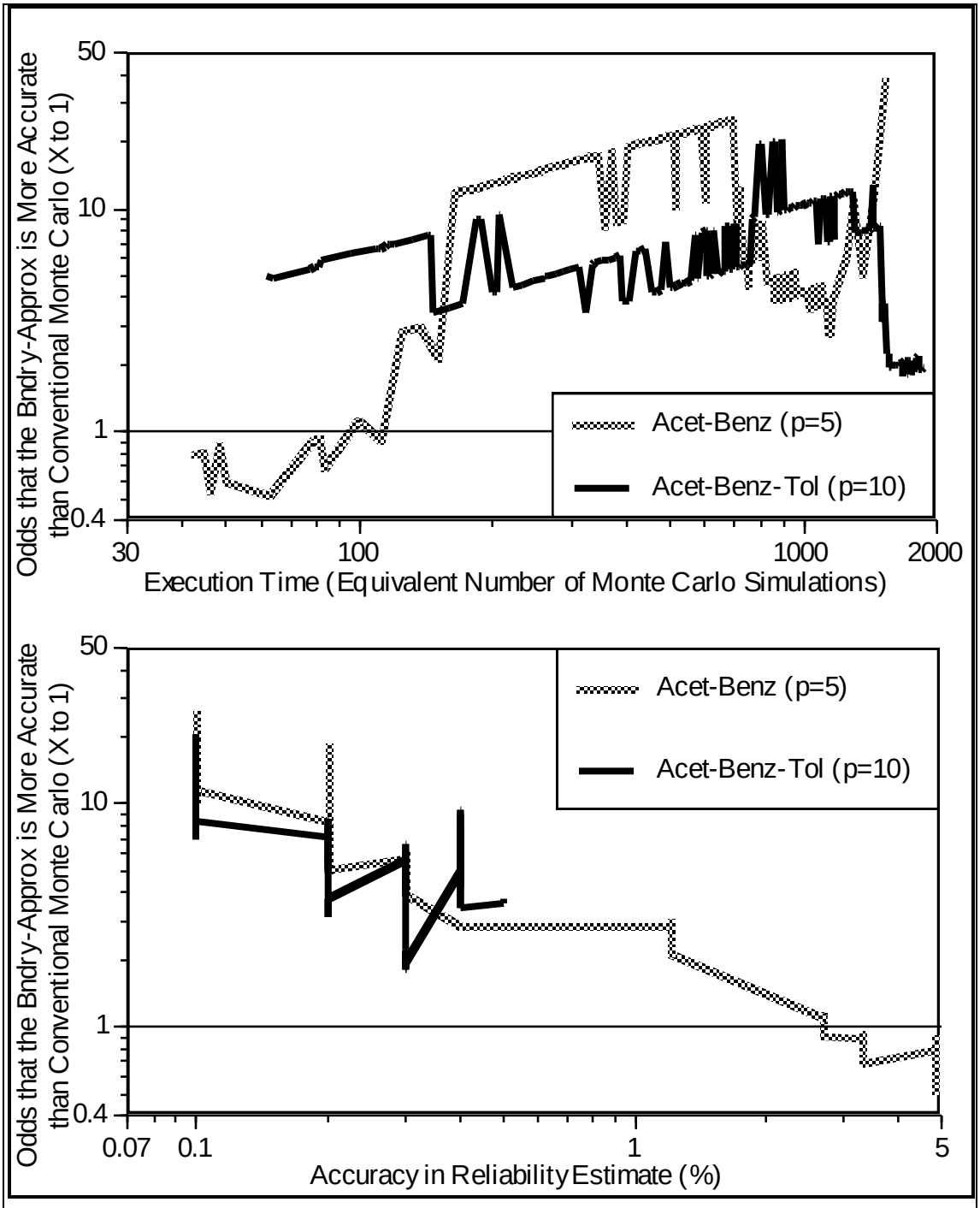


Figure E-2-4
Likelihood that the Boundary-Approximation Procedure is More Accurate than the Conventional Monte Carlo Procedure as a Function of Computational Effort and Accuracy: Acetone-Benzene + Toluene

Appendix F. Boundary-Approximation Algorithm Flowcharts

This section shows more detailed algorithm flowcharts for the boundary-approximation procedure.

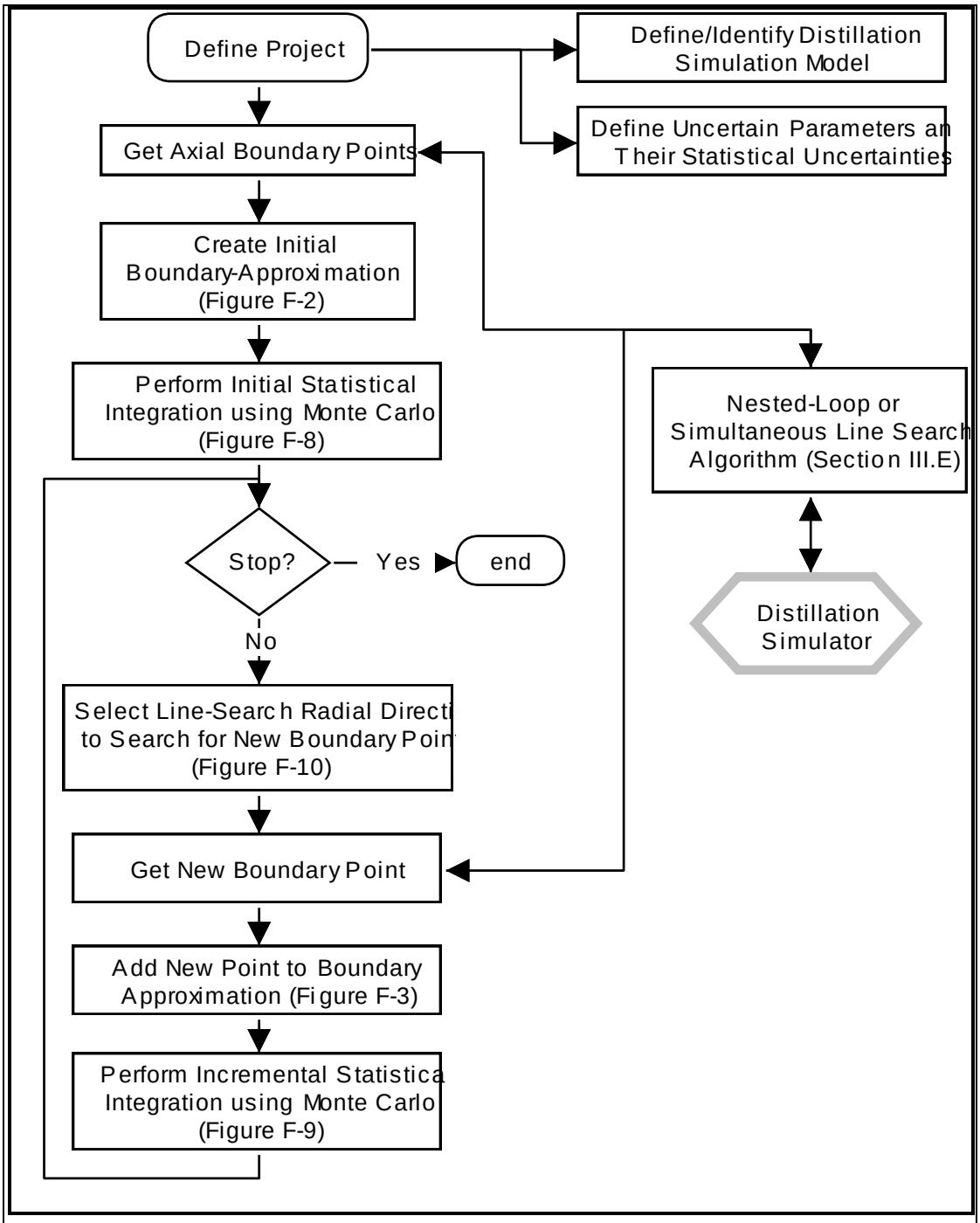


Figure F-1
Overall Convergence Algorithm for Boundary-Approximation Procedure

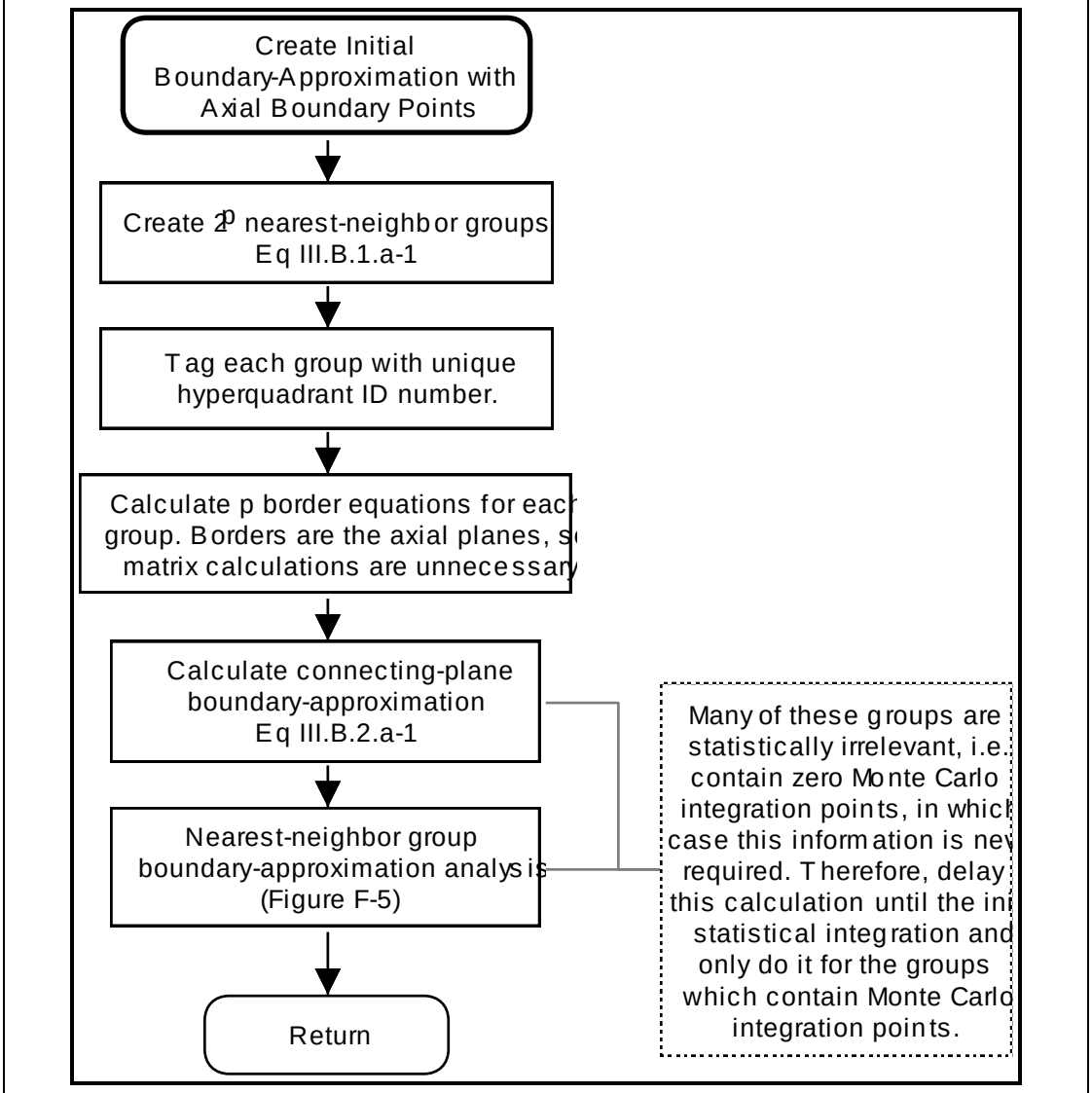


Figure F-2
Creation of Initial Boundary-Approximation With Axial Boundary Points

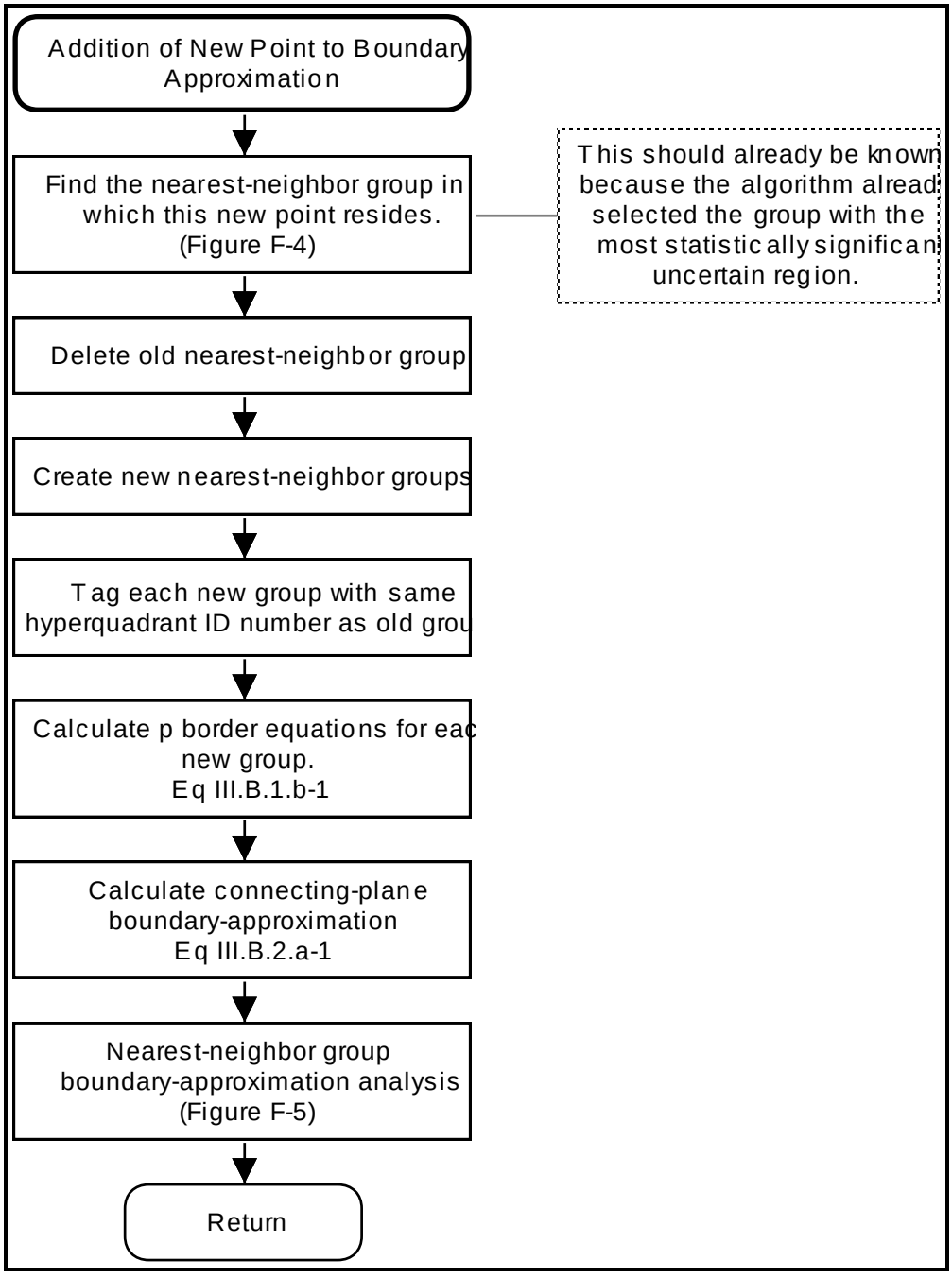


Figure F-3
Addition of New Boundary Point to Boundary-Approximation

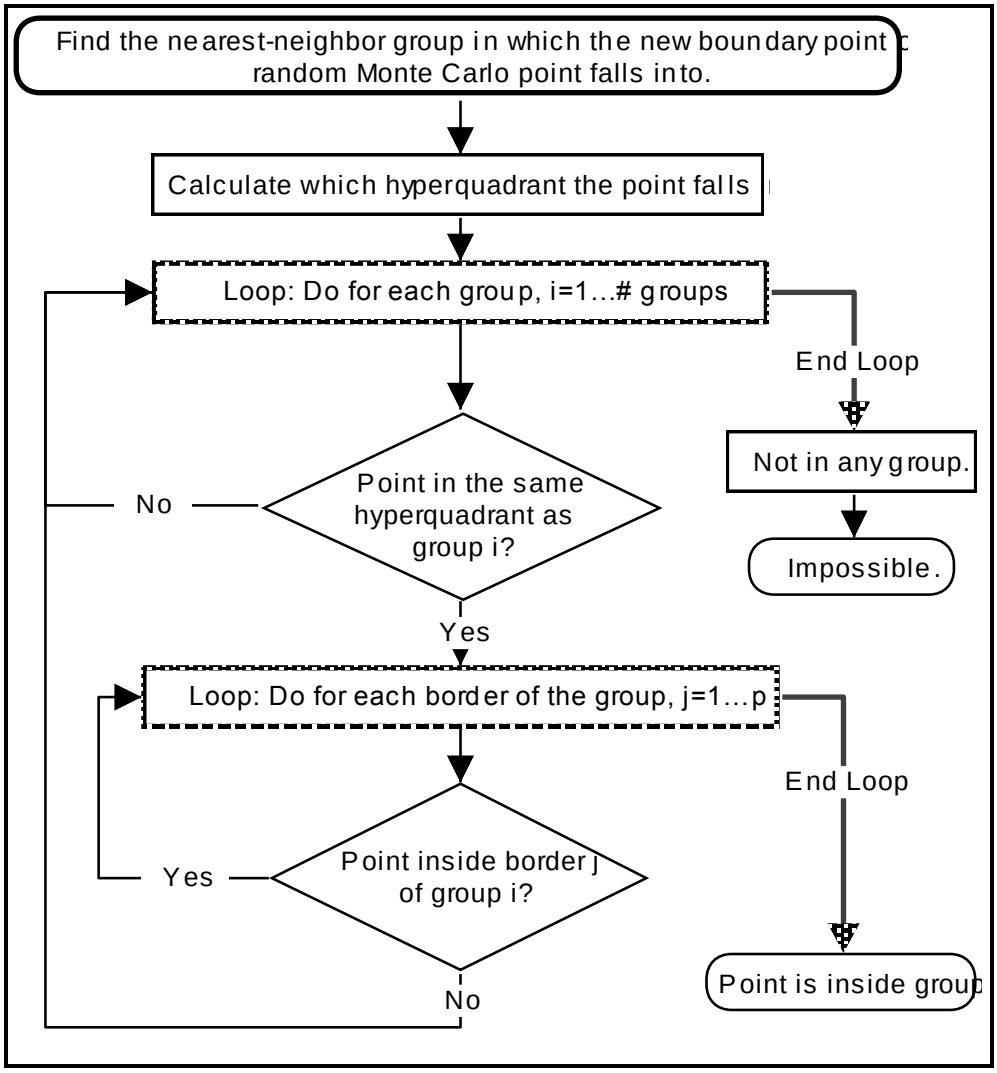


Figure F-4
Finding the Nearest-Neighbor Group in Which the New Boundary Point or Random Monte Carlo Point Falls Into

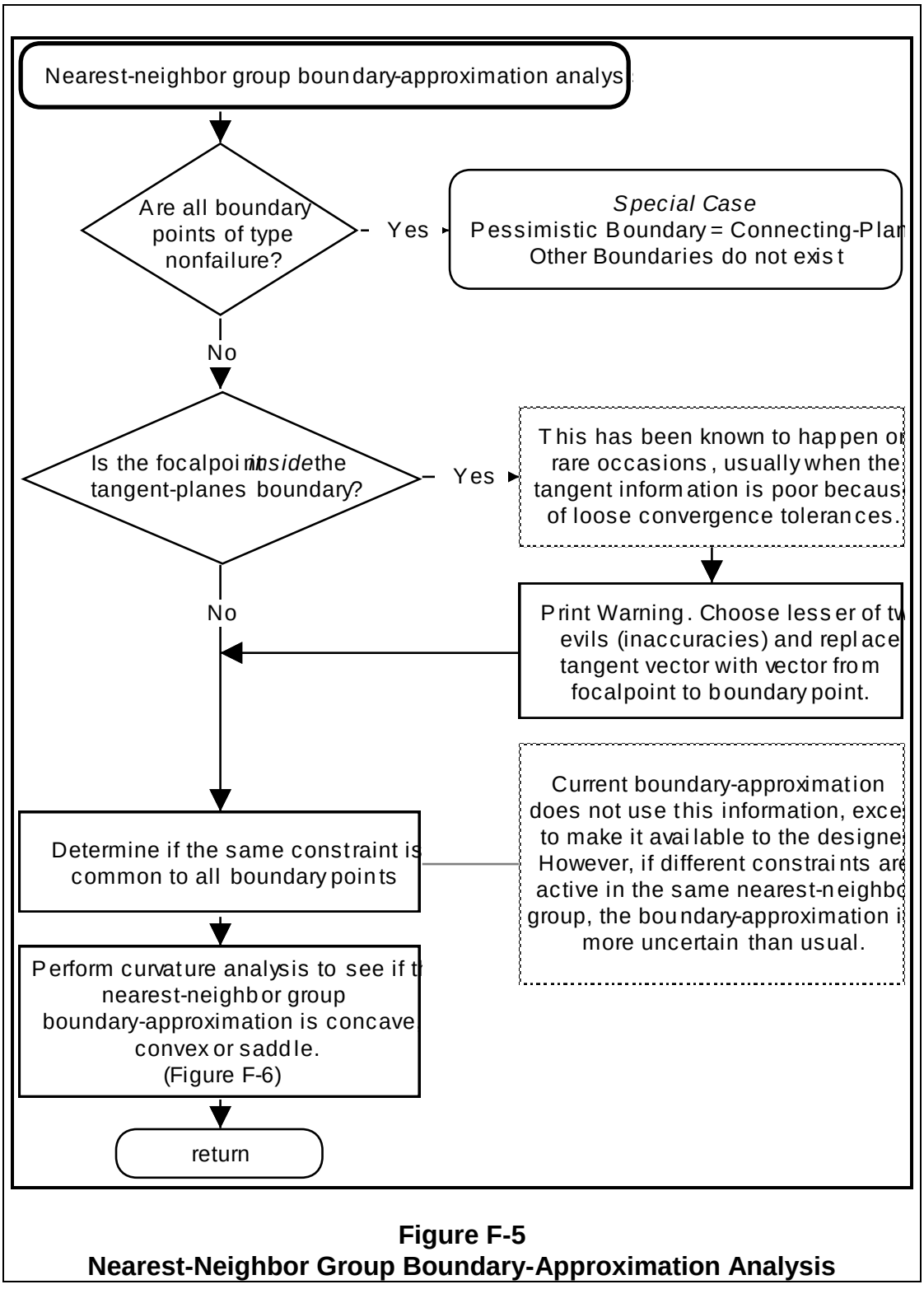


Figure F-5
Nearest-Neighbor Group Boundary-Approximation Analysis

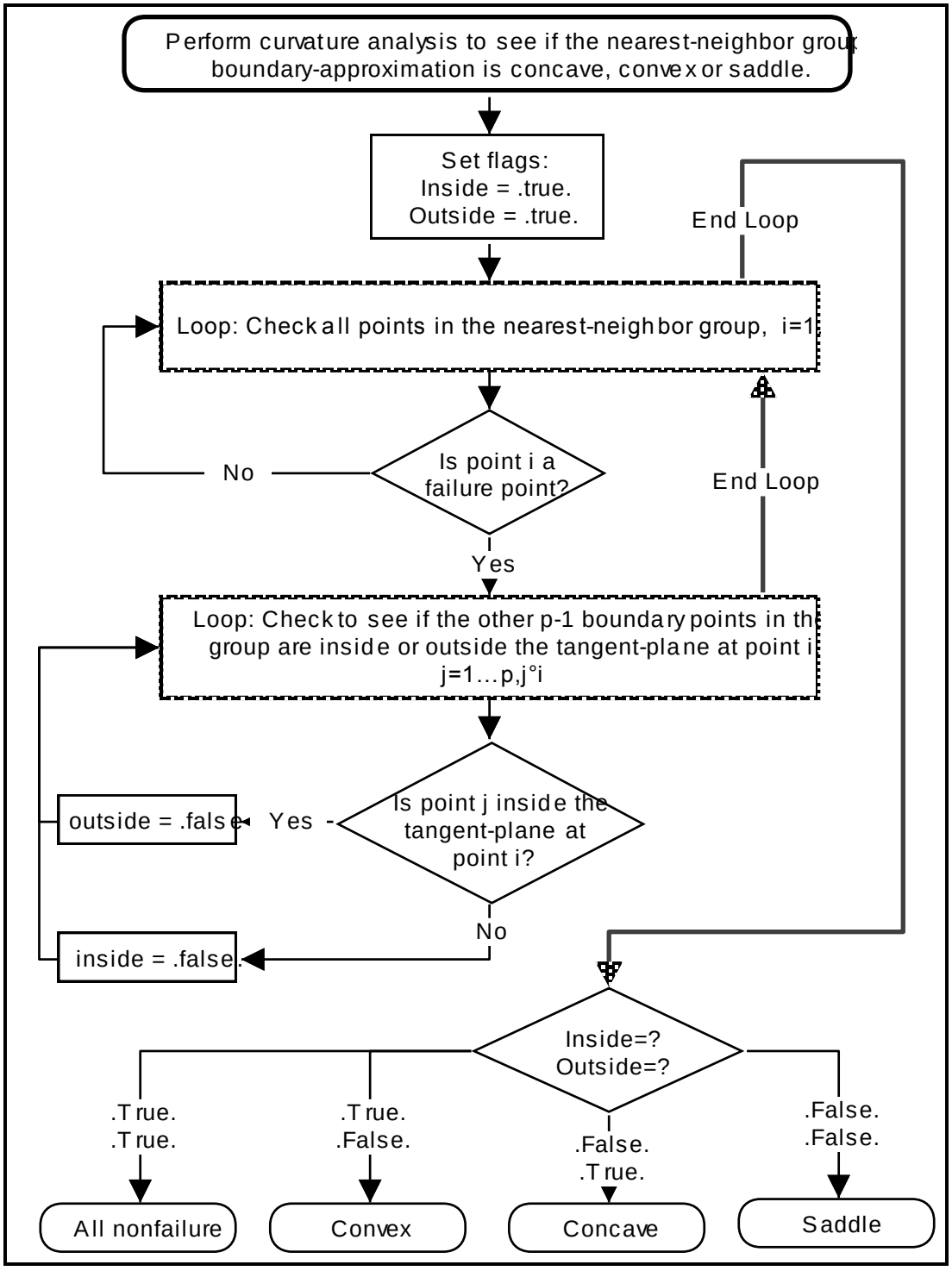


Figure F-6
Curvature Analysis of Nearest-Neighbor Group

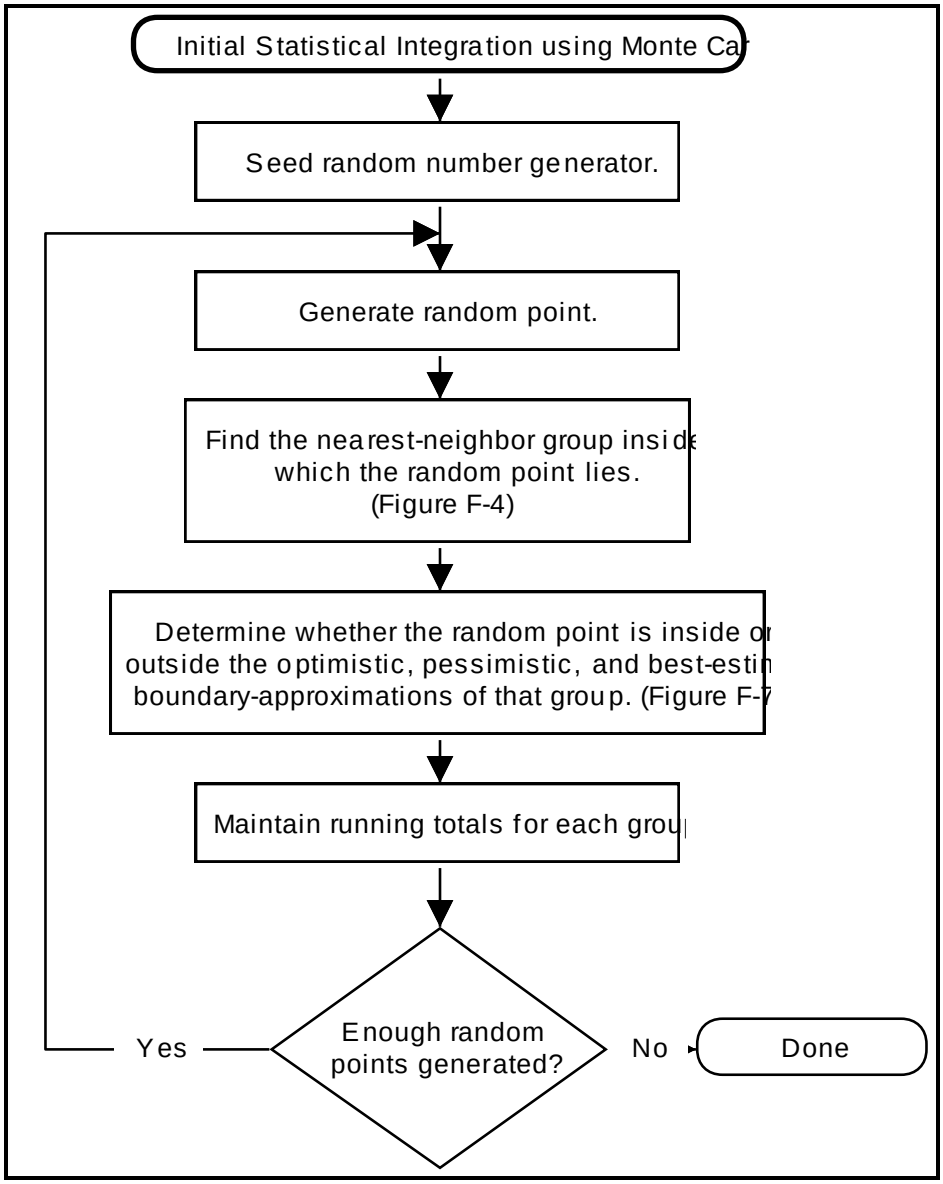


Figure F-8
Initial Integration Using Monte Carlo and Boundary-Approximation

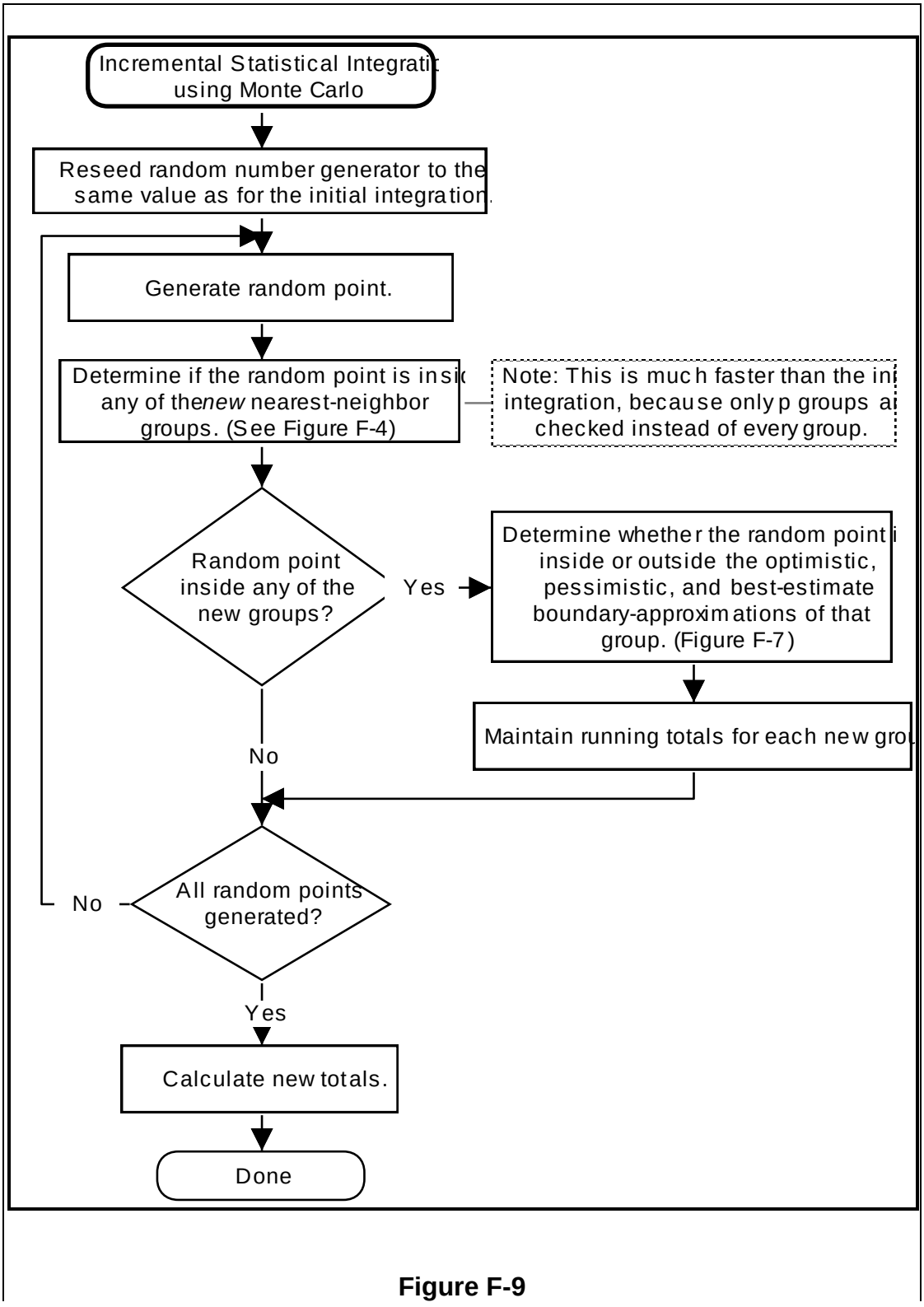

Incremental Statistical Integrati
using Monte Carlo
Reseed random number generator to the
same value as for the initial integration
Generate random point.
Determine if the random point is insi
any of the *new* nearest-neighbor
groups. (See Figure F-4)
Note: This is much faster than the ini
integration, because only p groups ai
checked instead of every group.
Random point
inside any of the
new groups?
Yes
Determine whether the random point i
inside or outside the optimistic,
pessimistic, and best-estimate
boundary-approximations of that
group. (Figure F-7)
Maintain running totals for each new gro
No
No
All random points
generated?
Yes
Calculate new totals.
Done

Figure F-9

Incremental Integration Using Monte Carlo and Boundary-Approximation

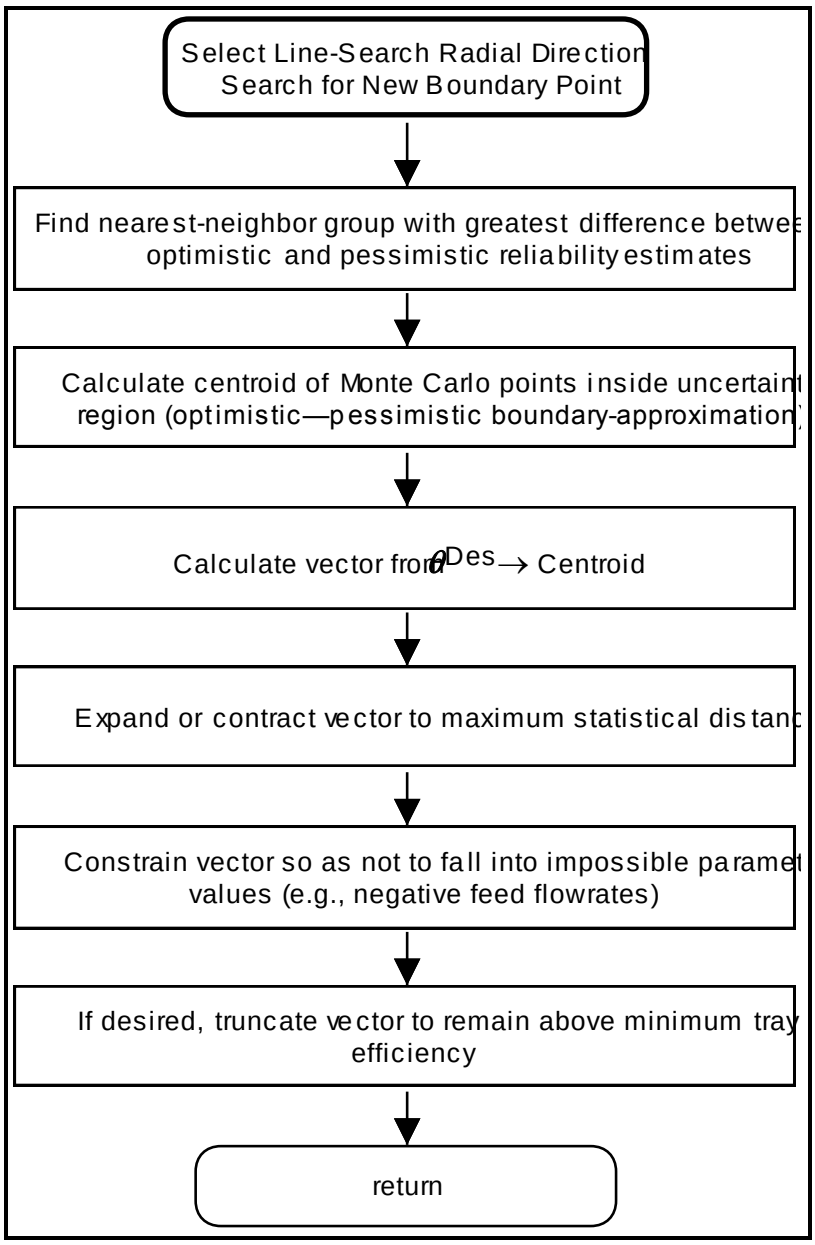


Figure F-10
Select Line-Search Radial Direction to Search for New Boundary Point

Appendix G. Sources of Errors in the Mathematical Boundary-Approximation Reliability Estimate

Section V.C.1.b. showed that the boundary-approximation may have slight errors in the reliability estimate even after convergence. This section discusses some possible causes.

Sources of error in the boundary-approximation reliability estimate can be divided into two types:

(1) Errors that may be minimized or eliminated by better computers and/or additional computational effort.

(2) Errors that are caused by inherent limitations of the boundary-approximation procedure and cannot be eliminated.

Examples of the former include computer numerical roundoff and inaccuracies in the boundary point+gradient information caused by loose convergence criteria. Examples of the latter include hidden regions caused by boundary curvature 'around the corner', or multiple constraint boundaries in a given radial direction. Naturally, the latter are of greatest concern. However, as discussed in this section, their likelihood of occurrence should be statistically small.

Each source of error is listed below, and discussed where significant.

G-1. Minimizable Errors

- *Less than an infinite number of boundary points from which to interpolate the boundary-approximation*

Since the true boundary is nonlinear, linear interpolation is inherently inaccurate. However, accuracy should improve as the number of boundary points increase. This error should be quantifiable without any hidden 'surprises'.

- *Limiting maximum radial line-search distances*

Limiting the maximum radial line-search distance (see section III.E.1.) may slow convergence by increasing the percentage of nonfailure boundary points. However, this is partially offset because each nonfailure point only requires one process simulation instead of several (see also section V.B.2.b). No systematic error is directly introduced by nonfailure points. However, secondary effects may cause conservative estimates of the best-estimate

and optimistic reliabilities, because of the increased number of nearest-neighbor groups containing only nonfailure points, for which the best-estimate and optimistic boundary-approximations are assumed not to exist (see section III.B.2.b). However, this is unlikely to be statistically significant unless the maximum radial distance is grossly short. Note: This error *was* once observed for a very high reliability test case (>95% reliability), with too short a radial distance. High reliability designs should use longer maximum radial distances.

- *Inaccurate constraint boundary point+gradient information, caused by loose convergence criteria for: (a) the distillation simulation model convergence, (b) line-search constraint boundary point convergence, and (c) gradient approximation by finite difference*

Loose convergence criteria cause errors in the constraint boundary point and gradient information. This in turn causes an inaccurate boundary-approximation. Tighter tolerances improve accuracy at the cost of greater computational effort. Computer precision imposes a practical accuracy limit. An optimum convergence tolerance exists, which provides accurate information for reasonable cost.

Since the distillation model is too complex for analytical derivatives, finite-differencing is generally the most convenient procedure. However, this can be inaccurate. If the step distance is too small, errors occur due to computer roundoff. If the step distance is too large, second order curvature becomes significant. However, error due to iterative convergence is likely to be more significant. This is discussed in section IV.B.4.

Slight inaccuracies in the point-gradient calculations can cause incorrect boundary curvature analyses, which in turn causes boundary-approximation error. This is most likely to occur when the constraint boundary is fairly linear and the boundary points are far apart—a small gradient error can mean the difference between concave, convex, and saddle curvatures.

- *Computer roundoff in the nearest-neighbor point grouping and border calculations*

Section III.B.1. describes the mathematics of organizing the constraint boundary - approximation into a set of nearest-neighbor groups. One requirement is a mathematical description of the borders between nearest-neighbor groups. For each group, the calculation requires p matrix inversions of order p. These matrices are sometimes nearly singular, so some computer roundoff may occur and cause slight inaccuracies in the borders between neighbor groups. The net effect is that for large dimensions and many boundary points, it is possible for a few of the randomly generated Monte Carlo points to ‘fall between the cracks’ of the group border separations. That is, a random Monte Carlo point might not be inside any

neighbor group which is an impossibility. Solutions include: (1) Increase the numerical precision; (2) Give the random point a small arbitrary perturbation and retry until it falls within one of the neighbor groups; (3) Use a more robust matrix inversion routine; (4) Ignore the random point and generate a new one instead. I chose the latter option, because it should not cause a large systematic error.

During testing with the distillation problem, this problem did occur, but was minor. For p=5, this only occurred once or twice every 1,000 Monte Carlo points. So the resultant error will be at most ±0.2% and probably much less because the error is not inherently systematic.

- *General Computer Roundoff*

The computer has roundoff errors due to limited numerical precision. However, this should be inconsequential in comparison with other errors.

G-2. Unavoidable Errors

- *A boundary that curves 'around the corner' as viewed from the focalpoint*

The boundary-approximation procedure cannot look 'around the corner' of the constraint boundary, as shown schematically in Figure G-2-1. Note that the boundary-approximation procedure cannot 'see' the shaded region. Errors may be positive or negative depending on which region the initial simulation point falls into.

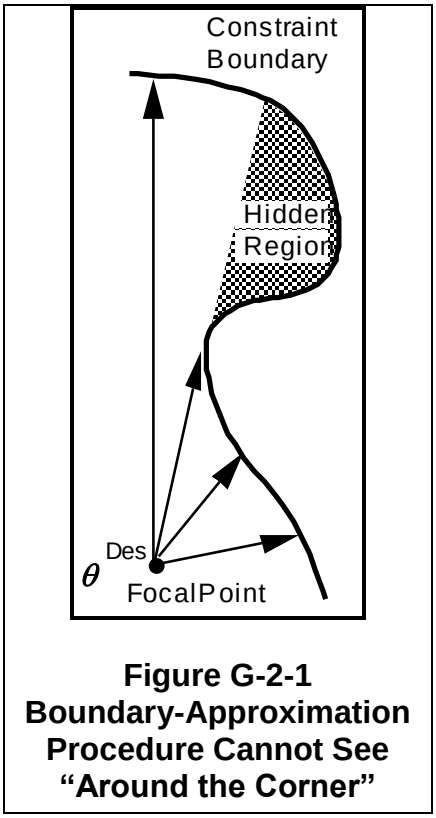


Figure G-2-1 Boundary-Approximation Procedure Cannot See "Around the Corner"

Except for parameter transformations to linearize the constraint boundary shapes, there appears to be no practical way of correcting this problem if it exists. While we hope that this situation is highly unlikely and statistically insignificant, exploratory studies for the various process operations should be performed to indicate if this curvature exists.

Distillation constraint boundaries *do* exhibit this behavior as seen in Figures C-1.d.ii-1, C-2.d.ii-1 and C-3.d.ii-1 for the database λ-space . Here, the curvature is 'banana-shaped'. However, note that this curvature occurs only in statistically insignificant regions and so is unimportant.

- *Multiple constraint boundary points in a radial direction as viewed from the focalpoint*

Multiple solutions to the radial line-search may also occur, as seen in Figure G-2-2. This occurred for both the 2MB1-IPM and Acetone-Benzene distillation test problems in the database λ-space. However, the second boundary point (2) occurred along the principal axis at about ±30 standard deviations from the design point, which is vanishingly insignificant. This situation is comparable to what happens when data are overfitted to a high order polynomial. While there may be good agreement within the fitted region, the curve will oscillate wildly in the extrapolation region.

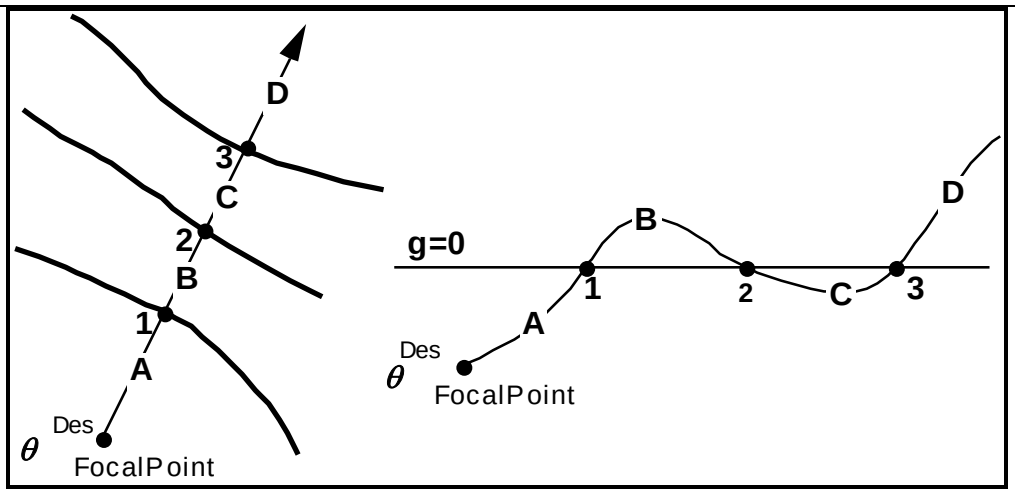


Figure G-2-2
Multiple "Waves" in Inequality Surfaces Causes Multiple Constraint Boundary Point Solutions

- *Boundary 'ripples' caused by a non-monotonic surface*

A more realistic source of error is undetected nonlinear 'ripples' in the constraint boundary. A 'ripple' can be defined as a region within a nearest-neighbor group where the boundary curvature is nonmonotonic. An example of this is shown schematically in Figure G-2-3. Judging from the point+gradient information at the two boundary points, the curvature is convex. So the tangent-planes boundary is the best-estimate and optimistic boundary, and the connecting-plane boundary is pessimistic. However, as shown the true boundary actually curves *inside* the connecting-plane boundary, which violates the pessimistic assumption.

If the nearest-neighbor group is statistically significant, this error will be detected by further investigation. However, if the uncertainty region is small, then this neighbor group will not be investigated further, and the error will not be corrected.

Figure G-2-3 describes an overestimation of the design reliability. It is also possible for this error to cause underestimation of the reliability.

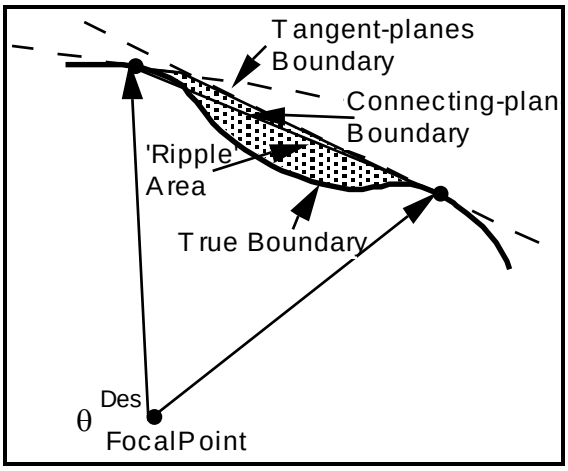


Figure G-2-3
Boundary 'Ripple" Causes Undetected Systematic Error

Appendix H. Program Listings

Did you hear the one about the programmer who got stuck in the shower indefinitely? He took the instructions on the shampoo bottle literally: “Lather. Rinse. Repeat.” MacUser, 1992.

This section lists the Fortran codes developed for this research. Appendix H-1 provides some flowcharts for the user menu and interface. Appendix H-2 lists the names of the programs and subroutines with a brief description. Appendix H-3 provides an alphabetical list of these routines, however, due to excessive length these routines are not printed, but are supplied on the accompanying floppy diskette.

H-1. User Interface Flowcharts (User Control Menus)

Figure H-1-1 provides a flowchart of the startup operation, which performs some initialization, then waits for the user to select further options from the menu. The menu options are listed in Figures H-1-2 to H-1-6.

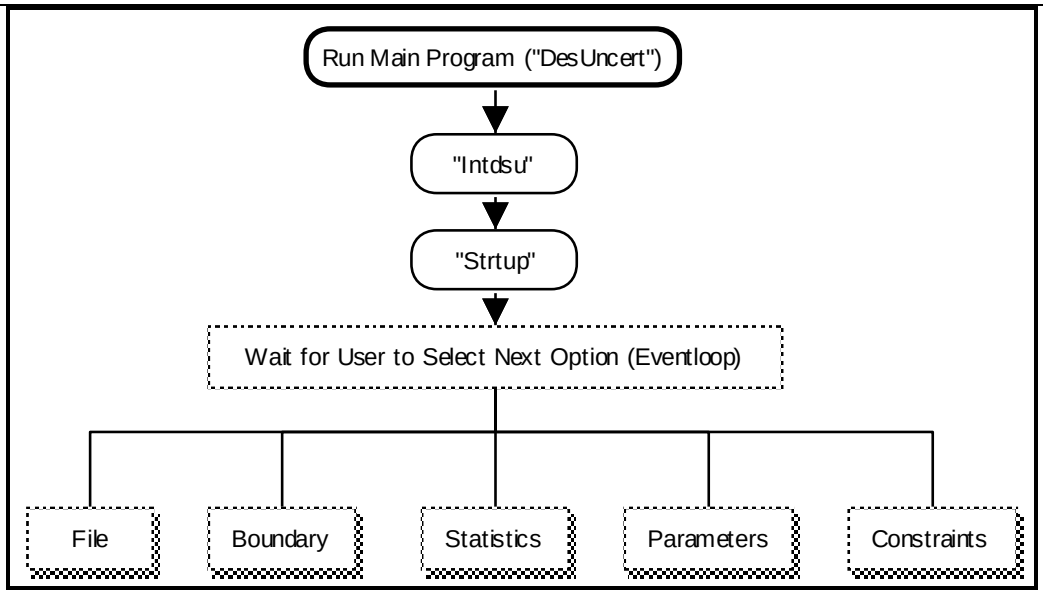


Figure H-1-1
Program Initialization and Eventloop

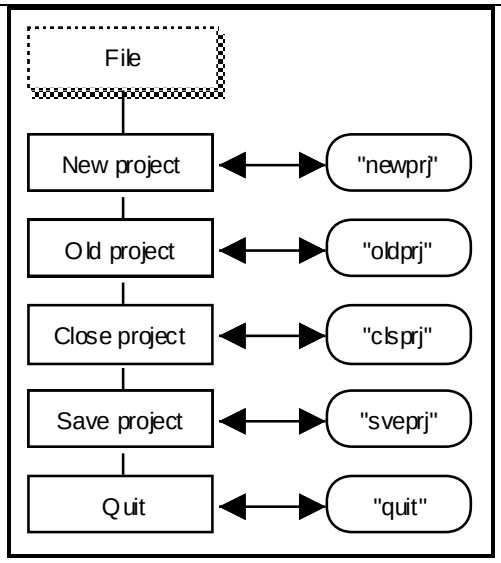


Figure H-1-2
File Menu

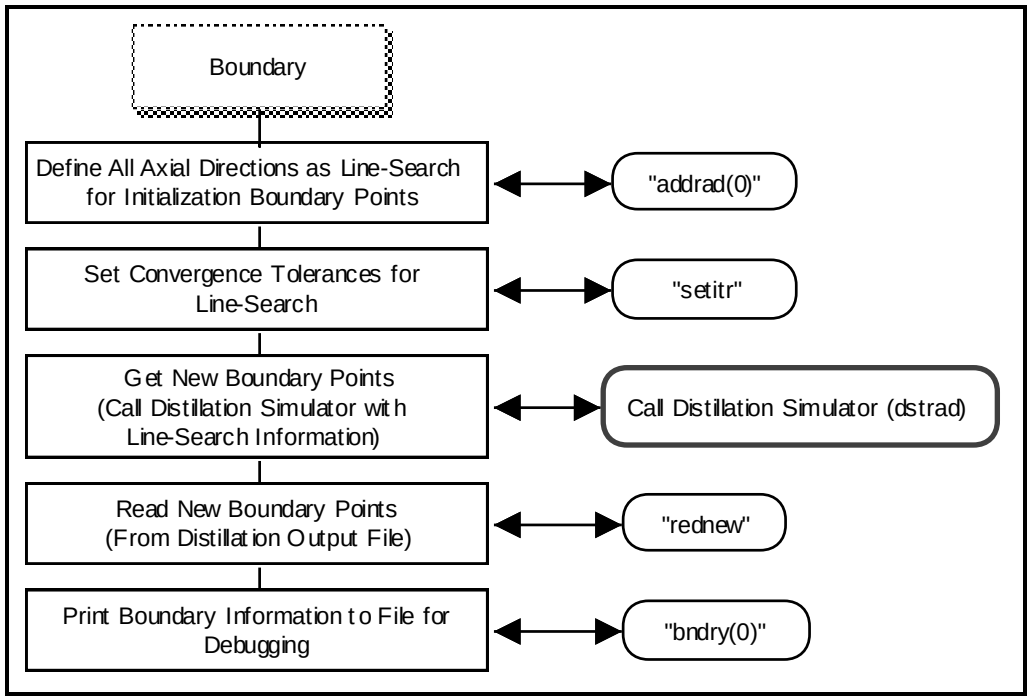


Figure H-1-3
Boundary Menu

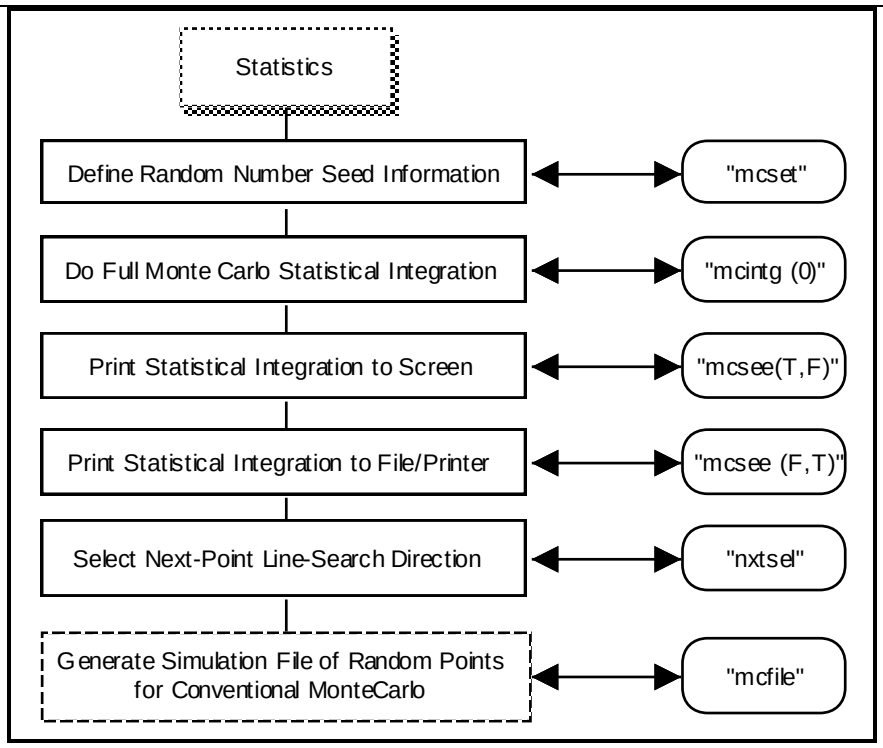


Figure H-1-4
Statistics Menu

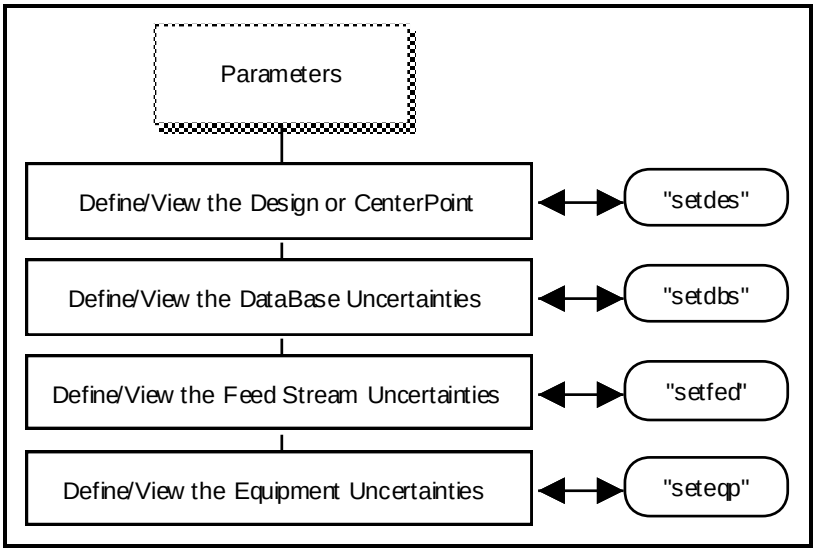


Figure H-1-5
Parameters Menu

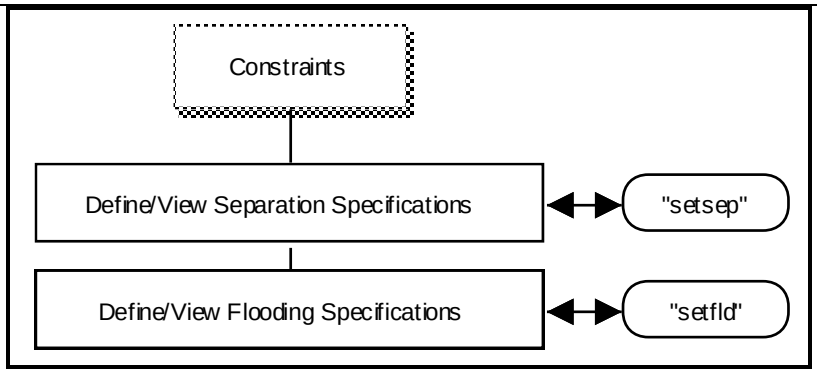


Figure H-1-6
Constraints Menu

H-2. Program and Subroutine Descriptions

This section lists the names of the Fortran program and subroutines for the project and provides a brief description of each.

Main Program

DesUncert Run control program for the boundary-approximation procedure. Program sets up the menubar, calls the initialization subroutine (intdsu) to initialize the variables, then sits in an eventloop waiting for the user to select some option from the menubar.

Global Declaration Files (Include Files)

MaxDimen Dimensions maximum number of parameters, constraints, boundary points and nearest-neighbor groups. This is because Fortran does not provide dynamic runtime arrays. This file allows dimensioning to be changed in a single location.

DimenCom Companion File for *MaxDimen* to declare the variables in common.

CompType Declares certain information for different computer systems, e.g., screen unit identifier.

CtrlDec Contains file project and storage pathname

CtrlCom Companion File for *CtrlDec* to declare the variables in common.

DesUncDec Declares variables for the design with uncertainty calculations, including boundary points, boundary-approximation mathematics, design reliability, etc.

DesUncCom Companion File for *DesUncCom* to declare the variables in common.

DistDec Declares variables for communication with the distillation simulator

DistCom Companion File for *DistDec* to declare the variables in common.

StatDec Declares variables for the statistical descriptions of the uncertainties

StatCom Companion File for *StatDec* to declare the variables in common.

Startup Subroutines

Intdsu Initializes all variables

Strtup Starts up a current project. Asks the user for the projectname, looks for control files, and reads in known boundary points.

Redprj Reads in the control file information to set up the boundary-approximation procedure

Redpts Reads in the boundary points that have already been obtained and that are stored in a data file

Project Control (File Menu) Subroutines

Newprj Creates a new project

Oldprj Reads an existing project into memory

Sveprj Save project information to file

Saveas Saves an old project as a new project with a different name

Clsprj Closes the current project

Prtprj Prints project information (not developed)

Quit Prompt to save any changes and quit

Svedef Save default settings for various information (not developed)

Boundary-Approximation Subroutines

Distillation Simulator Communication

Getpts Call distillation subroutine to find the boundary point given the line-search direction. Unfortunately, due to Fortran programs being unable to communicate with each other, this subroutine was not written. Instead, two separate Fortran programs (*DesUncert* and *Distillation*) were run simultaneously in the MacIntosh MultiFinder and a macro utility was used to run each program in sequence. Communication was accomplished by data files.

Rednew Read new boundary point from data file and fit into boundary-approximation (communication with distillation simulator).

Line-Search and Next-Point Selection Subroutines

Addrad(0,1,2)	(0) Called at initialization to write a data file setting up the axial line-search directions to create the initial boundary point set. (1) Create a new data file containing the next-point line-search direction. (2) Append next-point line-search to data file (allows more than one new boundary point at a time between Monte Carlo integrations).
Nxtrad	Determine next-point line-search direction
Sdvdst	Calculate the standard deviation distance of the line-search vector (used to normalize search distance)
Setitr	Set convergence tolerances for line-search procedure

Mathematics of Nearest-Neighbor Groupings and Tangent-Plane and Connecting-Plane Boundary-Approximations

Bndry	Controlling subroutine to call other subroutines
Inconn	Determine if Monte Carlo point is inside or outside the connecting-plane boundary-approximation of the nearest-neighbor group
Intan	Determine if Monte Carlo point is inside or outside the tangent-plane boundary-approximation of the nearest-neighbor group
Neigha	Add new boundary point to boundary-approximation
Neighb	Calculate border-planes for nearest-neighbor group
Neighd	Calculates ID numbers of nearest-neighbor groups from hyperquadrant location
Neighi	Initializes boundary-approximation
Neighp	Create connecting-plane boundary for nearest-neighbor group
Neighs	Determine if nearest-neighbor group has the same active constraint at each of its boundary points
Neight	Determine boundary curvature for the nearest-neighbor group
Neighw	Find which nearest-neighbor group the new point is in
Tanclc	Determine if Monte Carlo point is inside or outside the tangent-plane of a particular boundary point

Statistics Subroutines

MCFile	Generate data file containing list of Monte Carlo random points for distillation simulator to read in and calculate–Conventional Monte Carlo calculation
MCGen	Generates the next Monte Carlo random point

MCIntg	Perform Monte Carlo integration with the boundary-approximation procedure
MCSee	Print statistical integration information to screen or data file
MCSet	Define Monte Carlo integration information and seed random number generator
NxtSel	Select type of next-point line-search direction technique (midpoint; centroid)
Rand	Generates a uniformly distributed random number
Randnm	Generates a normally distributed random number

Parameters Menu Subroutines

SetDbs	View/Change the database parameter uncertainties
SetEqp	View/Change the equipment parameter uncertainties
SetFed	View/Change the feed stream parameter uncertainties

Generic Mathematical Subroutines

Matmul	Multiply two matrices together
Matslv	Perform a matrix inversion (solve simultaneous equations)
Str255	Converts Fortran string to Pascal Lstring (used for MacIntosh file and projectnames)
Zero	Zeroes an array

Line-Search Algorithms

Distillation Nested-Loop Algorithm

DstRad	Distillation subroutine to perform the line-search. Appears in menu of distillation program and is selected by a MacIntosh operating system macro utility.
DstRrr	Distillation subroutine to perform the Conventional Monte Carlo simulation. Reads in data file generated by *MCFile*, solves the simulation for each point, and write an output file.

Simultaneous Line-Search Algorithm

The simultaneous algorithm was developed and studied using geometric test problems. When control variables are not present, the algorithm was satisfactorily confirmed. However, when control variables are explicitly present, the algorithm was only partially tested and still requires further work. When testing was switched to distillation, the simultaneous line-search algorithm was put aside. The following files list the code for the simultaneous line-search. However, it may not be fully debugged.

BndRad	Controlling subroutine to perform the simultaneous line-search algorithm
Dampsb	Allows user to set damping factors for convergence algorithm
Deltth	Performs the iteration step in θ (z and x are unchanged)
Deltx	Performs the iteration step in x only (θ and z are unchanged)
Deltxt	Performs the iteration step in x given the step in θ (z is unchanged)
dgdtsb	Calculates $\partial \mathbf{g} / \partial \theta$
dgdxsb	Calculates $\partial \mathbf{g} / \partial x$
dgdzsb	Calculates $\partial \mathbf{g} / \partial z$
dhdtsb	Calculates $\partial \mathbf{h} / \partial \theta$
dhdxsb	Calculates $\partial \mathbf{h} / \partial x$
dhdzsb	Calculates $\partial \mathbf{h} / \partial z$
Dotprd	Calculates angle between radial direction vector and $\partial \mathbf{g} / \partial \theta$. At convergence, angle is zero, so this is a useful for tracking convergence.
Gcalc	Calculates $\mathbf{g}$
Hcalc	Calculates $\mathbf{h}$
Logicc	User specifies convergence logic
MatInv	Performs forward gaussian elimination for step in θ
Orthcl	Calculates orthogonal boundary points for each inequality constraint
SumSq	Calculates sum of squares adjustment in theta for convergence tracking purposes (minimized when finding orthogonal boundary points in *Orthcl*)
Tgradc	Calculates orthogonal vector to constraint boundary ($\partial \mathbf{g} / \partial \theta$ such that $\mathbf{h} = \mathbf{0}$)
Zgradc	Calculates steepest descent vector in z
Zoptim	Performs search along steepest descent vector to minimize $\mathbf{g}$ with respect to z

H-3. Program and Subroutine Listings

Due to excessive length, the files are not printed, but provided on a separate accompanying MacIntosh floppy diskette.

Appendix I. Random Number Generator

This section documents the MacIntosh pseudo-random number generator.

The MacIntosh pseudo-random number generator was used for all Monte Carlo calculations. The "randomness" of the generator is not critical for this work, however some basic tests for randomness were satisfactorily performed in Appendix B-1. This section presents documentation of the generator at the request of the committee.

The MacIntosh routine uses the following equation (Goldberg, 1991):

$$u^{k+1} = \mathbf{a}u^{k} \bmod \mathbf{b} \qquad \text{Eq I-1}$$

where

u is a pseudo-random number

$\mathbf{a} = 7^5 = 16807$

$\mathbf{b} = 2^{31} - 1 = 2147483647$

mod is the modulo (remainder) function

This is called the "prime modulus multiplicative linear congruential generator" or less formally the Lehmer generator (Park and Miller, 1988). It is a full period generating function (period = 2^{31}), is demonstrably random, and can be implemented correctly on almost any computer system. Note that all initial seeds between 1 and 2^{31} are equally valid. Park and Miller (1988) recommend it as the standard against which all others should be judged.

Appendix J. A Better Nearest-Neighbor Grouping Method

This section was developed after the thesis was completed. Watson's (1988) grouping method avoids the problems of my nearest-neighbor grouping approach and is recommended for further study. Watson's (1988) method allows overlap between nearest-neighbor groups, but this can be handled simply by using only the group that is closest.

The separation of the boundary-approximation into nearest-neighbor groups was handled by generating border-planes between all groups (section III.B.1.b.) This requires a substantial amount of computer memory and/or hard disk storage. More significantly, it also leads to inferior boundary-approximations as discussed in section III.B.1.c, namely that boundary information is not shared between hyperquadrants, the best set of nearest-neighbor groupings is not always generated (smaller groupings are skipped, and 'sliver-like' or highly acute angled groupings may result), and boundary points falling exactly on a border-plane are not properly handled.

Watson's (1988) grouping method avoids these problems, but allows overlap between neighbor groups. This is not a problem when adding new boundary points to the boundary-approximation, but it may cause inconsistency for the Monte Carlo integration. Specifically, if a random Monte Carlo point falls inside two or more neighbor groups, what should be done if it is inside one group's best-estimate boundary but outside another's?

This problem should be minor compared to the advantages. A simple solution is to use the boundary-approximation of the *closest* neighbor group. Therefore, I recommend study of Watson's grouping approach to replace the border-planes approach.

Watson's method requires calculation of each group's circumcenter and circumradius. A new point is inside the group if and only if the distance between it and the circumsphere is less than the circumradius. Unfortunately, Watson does not explain how to calculate the circumsphere (the circumradius may be easily calculated afterwards, as it is the distance from the circumsphere to any of the group's points). The circumcenter is *not* generally the centroid of the group's points. After much puzzling, I am still unable to find a simple solution. However, I have been able to find a different approach that may be simpler and does not require the circumcenter. Simply normalize all points to the unit hypersphere, fit a hyperplane through these points, then check new points (after normalization) to see on which side of the hyperplane they fall. This is fast and requires little memory. More specifically:

1) Normalize all boundary points to the unit hypersphere, i.e., place the origin at $\boldsymbol{\theta}^{Des}$, calculate the vector from $\boldsymbol{\theta}^{Des}$ to the boundary point, calculate the sum-of-squares of the vector's coordinates, then divide each coordinate by the square root of this value (the vector now has unit length).

2) Calculate the hyperplane connecting the normalized p points in each neighbor group, using Eq III.B.2.a-1 ($\mathbf{B} \cdot \mathbf{b} + \mathbf{1} = 0$).

3) To determine if a new point ($\boldsymbol{\theta}^{Sim}$) is inside the neighbor group, first normalize the point as described above, then insert it into the equation $\mathbf{b}^T \cdot \boldsymbol{\theta}^{Sim} + \mathbf{1} = ?$ to determine the sign. If the result is positive, $\boldsymbol{\theta}^{Sim}$ is outside the group. If the result is negative, $\boldsymbol{\theta}^{Sim}$ is inside the group.

This method fits a hyperplane to pass through the group's p boundary points on a unit hypersphere. The hyperplane intersects the hypersphere to create a (p-1) dimensional hypersphere (e.g., a plane intersecting a sphere creates a circle). If a new point is inside the group, it is on the "outside" of this intersection, i.e., further from the origin.

Note: Watson's method creates the set of *natural neighbor* groups. Thus, this Nomenclature should replace my *nearest-neighbor* groups.

Appendix K. Conventional Monte Carlo Variance-Reduction Techniques

Several modifications to Conventional Monte Carlo have been proposed to reduce its imprecision. If applicable, these would reduce the advantage of the boundary-approximation procedure. This section examines some of these modifications and presents some numerical studies. It is felt these modifications, while useful at low dimensions, will become cumbersome and inefficient at higher dimensions, and are not competitive with the boundary-approximation procedure.

This research compared the boundary-approximation procedure to Conventional Monte Carlo integration, as developed in sections III.D.2.a. and Appendix B. However, there are alternative numerical integration techniques which are similar to Monte Carlo but with potentially narrower confidence intervals. The conclusions of the research may therefore have been overstated because the boundary-approximation procedure was compared to an inefficient procedure (Conventional Monte Carlo).

However, these methods will not overthrow the boundary-approximation procedure as the best approach, because:

1) They may not automatically provide sensitivity analyses
2) They still require a significant number of simulation points to reduce the confidence interval to a satisfactory level
3) They all become inefficient at high dimensions
4) They are not extendible to hybrid statistics

This section presents a brief analysis of these variance-reduction techniques. The following is a list of alternative procedures that were found.

1) Hit-And-Miss Monte Carlo
2) Sample-Mean Monte Carlo
3) Control-Variate Method
4) Antithetic Variate Method
5) Quasi-Monte Carlo (Number-Theoretic) Methods
6) Stratified-Sampling Method
7) Bernstein or Orthogonal Polynomials and Product Rule Methods

1) Hit-And-Miss Monte Carlo

The Hit-and-Miss Monte Carlo method is the most primitive and slowest approach. See Yakowitz (1977) for details. This method is always inferior to Sample-Mean Monte Carlo, discussed next.

2) Sample-Mean Monte Carlo

This is the standard method most similar to that used in this research. The difference is that sample-mean Monte Carlo is primarily designed to integrate non-probability functions and uses *uniform* random distributions over the region of integration. Say we wish to evaluate $\int_0^1 f(\upsilon)d\upsilon$, where f is an arbitrary function and υ is an arbitrary integration variable. We generate a set of random values τ_i from a uniform distribution over [0,1], then calculate $\sum_{i=1}^{M} f(\tau_i)$, where M is the number of random points. The confidence interval is estimated from the sample variance. If the interval of integration is over a different interval, a variable transformation can be performed (Yakowitz,1977)[1].

This approach cannot be used if the region of integration is infinite. However, any integral of interest will go to zero at infinity, so the interval can be truncated. For example, a standard normal probability distribution can be integrated over [-5,+5] for negligible error.

However, a uniform distribution over [-5,+5] for the standard normal distribution will give a large number of near-zero function evaluations, which might be inefficient. This was tested with a one-dimensional standard normal distribution truncated to zero outside ±1.96 standard deviations (sdev), i.e., a 95% reliable “design”. This was integrated with the standard procedure used in this research (Conventional Monte Carlo) and a uniform random distribution using both ±5 sdev and ±2 sdev intervals. Results are shown in Figure K-1.

The Conventional Monte Carlo procedure was superior to the uniform distribution ±2 sdev run, which in turn was superior to the uniform distribution ±5 sdev run. This superiority was large initially and decreased with the number of random points. At 100 random points, the Conventional Monte Carlo had an error of 2%, the uniform/±2 sdev run had an error of 6.8%, and the uniform/±5 sdev run had an error of 24.7%. Confidence intervals appear to be accurate. Thus, we may conclude that the sample-mean approach should use as tight integration intervals as possible, but is inferior to the Conventional Monte Carlo procedure used in this research.

---

1 References listed at the end of this Appendix.

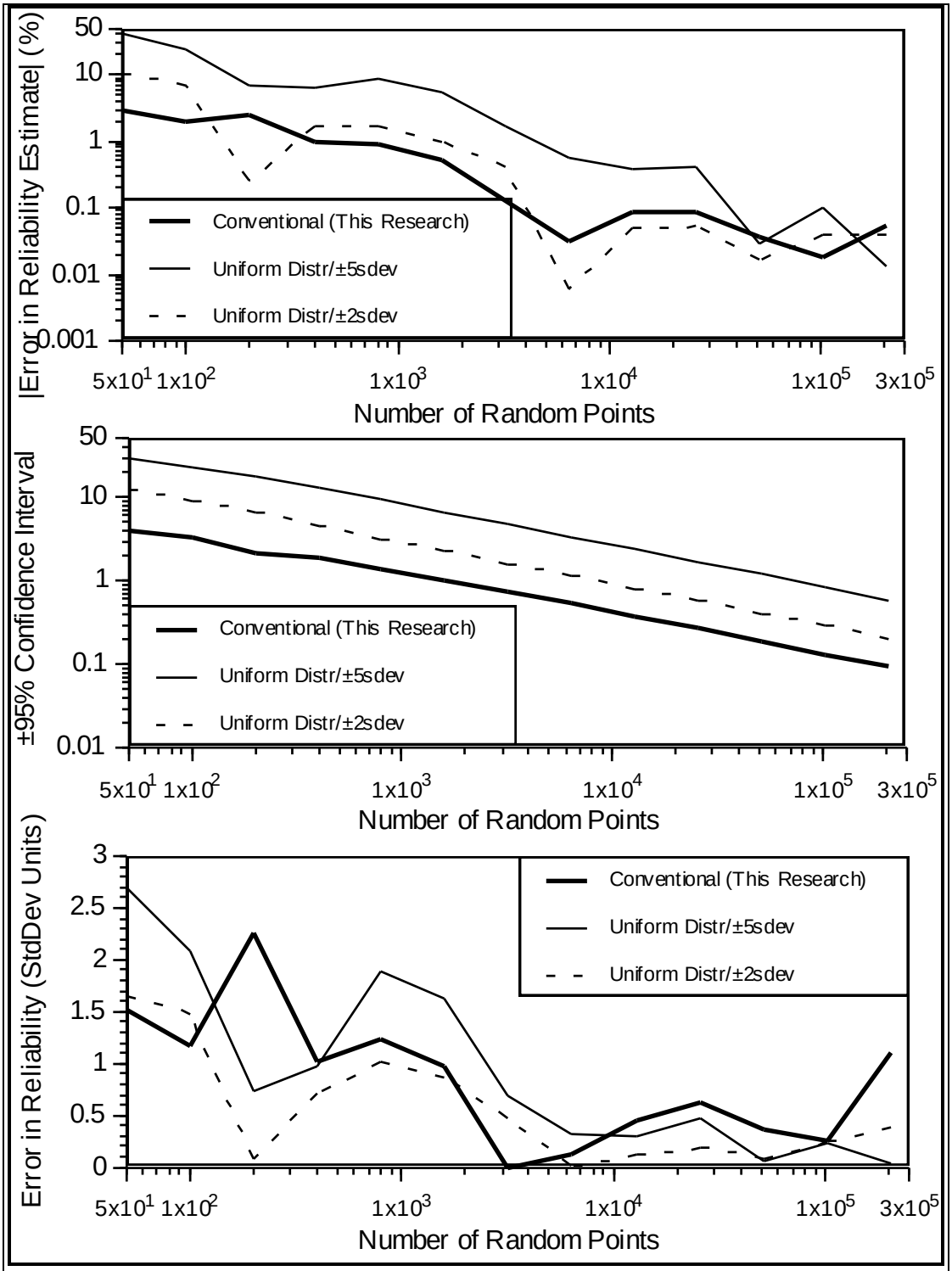


Figure K-1
Monte Carlo Integration of a 95% Reliability One-Dimensional Normal Distribution With Normal and Uniform Random Distributions

3) Control-Variate Method

The control-variate method depends on finding a function very similar to the integrand for which the analytical integration value is known. Then, the procedure is modified to integrate the *difference* between the two functions. Formally:

$$\int f(\upsilon)d\upsilon \;=\; \bar{F} + \int \left[f(\upsilon) - \bar{f}(\upsilon)\right]d\upsilon \qquad \text{Eq K-1}$$

where

$\bar{f}(\upsilon)$ is an analytically-integrable function as similar as possible to $f(\upsilon)$

$\bar{F}$ is the easily-found integral of $\bar{f}(\upsilon)$

Now the sample-mean Monte Carlo method is applied to the function $f(\upsilon) - \bar{f}(\upsilon)$ using a uniform random distribution. Since this function is smaller in value than $f(\upsilon)$, the variance is smaller.

One choice for $\bar{f}(\upsilon)$ is to set it to the probability distribution of the design parameters without the zero truncation at the constraint boundary. Therefore, the difference function is zero inside the constraint boundary and the value of the probability distribution outside the constraint boundary, which is an estimator for 1–←. This was tested for the same problem as before and is shown in Figure K-2 (Control-Variate / No Constraint Bndry). There is an improvement over Conventional Monte Carlo, shown by a significantly reduced error for the first two thousand random points, and a slightly tighter confidence interval.

Another option is to use an approximating polynomial or a step-function for $\bar{f}(\upsilon)$. To test this for the same problem, I used a linear interpolation function between ±2 sdev with 5, 40, and 100 evenly-spaced intervals. Results are shown in Figure K-2. There is a very significant improvement over Conventional Monte Carlo. Accuracy is about 10 times greater and correctly reflected in tighter confidence intervals.

Naturally, the more interpolation points, the better the results. However, this reaches diminishing returns—5 points are insufficient and 40 points is probably more than enough.

Extension to multiple dimensions will be very expensive. Even using only 5 points per dimension requires p^5 initial interpolation points—rapidly a prohibitively large number.

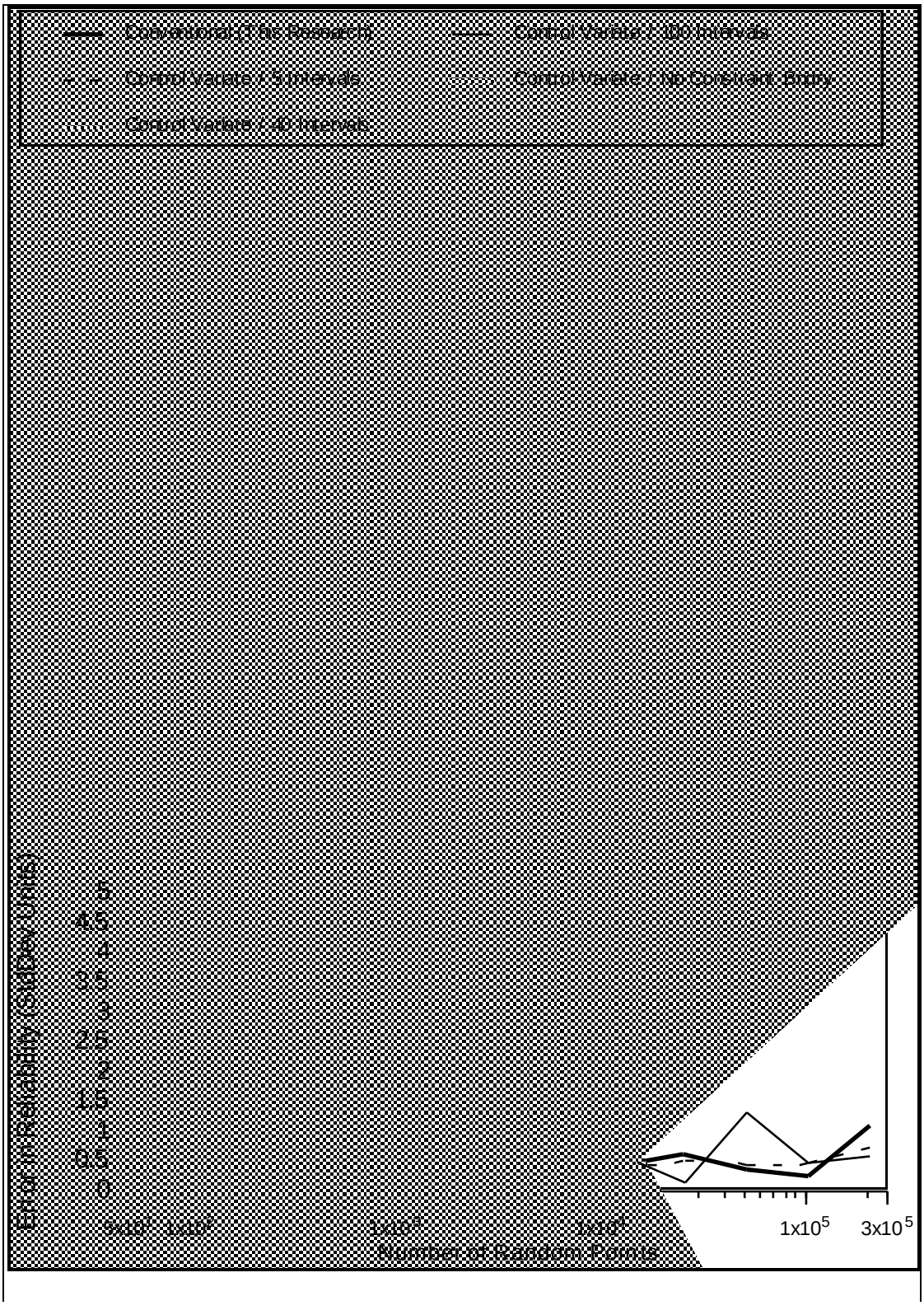

Conventional Monte Carlo
Control Variate + 90% Intervals
Control Variate + 95% Intervals
Control Variate + No Constraint Entry
Control Variate + 99% Intervals
Error in Option Price (Log Scale)
4.5
3.5
2.5
1.5
0.5
3x10^4
1x10^5
3x10^5
Number of Random Points

Figure K-2
Comparison of Conventional Monte Carlo and Control-Variate Methods

4) Antithetic Variate Method

This method depends on finding two different unbiased methods of estimating the integral. Then if these two methods are negatively correlated, the average of the two will have lower variance than either. While this is a well-developed method, Yakowitz (1977) notes that extension to multidimensions has proven difficult.

5) Quasi-Monte Carlo (Number-Theoretic) Methods

For these methods, the generated numbers are not random but are selected by rather complicated mathematics. Error is on the order of $O(\ln M/M)$ if the first-order derivative is continuous (not true for this research problem) and should be less than standard methods. Deák (1988) states that these methods have been successful for $2 \leq p \leq 10$, but will run into difficulties at high dimensions. They are also limited to rectangular regions of integration.

6) Stratified-Sampling Method

This method is described by Stroud (1971). The integration region is subdivided into several equal sub-regions, then one or more random and/or nonrandom points are selected within each sub-region. This always has lower confidence intervals than standard methods. This method will prove cumbersome at high dimensions, however, because of the very large number of sub-regions (Deák, 1988). Also, the amount of improvement may be small (Stroud, 1971).

7) Bernstein or Orthogonal Polynomials and Product Rule Methods

These methods generate random or nonrandom points and fit mathematical functions to these points to approximate the multidimensional function. They will become inefficient at higher dimensions (Deák, 1988).

High Dimensions

Deák (1988) reviews the state of the art and recommends standard Monte Carlo methods for dimensions greater than about five. All other methods will become complicated and inefficient. One problem is that the domain of integration becomes difficult at dimensions greater than one. Another significant problem is that the more complicated methods do not provide confidence intervals—error must be estimated by performing the calculation two different ways and using the discrepancy, which usually is an overestimation by an order of magnitude. Only standard Monte Carlo methods provide simple and accurate confidence intervals.

Overall Conclusion

All these methods provide only minor improvements over Conventional Monte Carlo, share many of the disadvantages, and require significant preparatory effort for each problem. This, plus their increasing complexity with dimension, makes it unlikely that they will find common usage in Chemical Engineering to replace Conventional Monte Carlo. The boundary-approximation procedure should not be challenged by these methods.

References for this section

Deák, I., 1988. "Multidimensional Integration and Stochastic Programming". *Numerical Techniques for Stochastic Optimization*, (Ermoliev, Yu, Wets, R. J-B Eds). Springer-Verlag, New York. 187-200.

Strong, A. H., 1971. *Approximate Calculation of Multiple Integrals*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Yakowitz, S. J., 1977. *Computational Probability and Simulation*, Addison-Wesley Publishing Company, Reading, Massachusetts.

Appendix L. Conventional Monte Carlo *Can* Perform Inexpensive Sensitivity Analyses

In this report I have criticized Conventional Monte Carlo for its inability to provide sensitivity analyses without repeating the integration, e.g., section II.D.2.a. This was incorrect. It is shown in this section how Conventional Monte Carlo *can* provide sensitivity analyses without any additional computational effort. However, the accuracy of this approach remains to be studied.

This section shows how Conventional Monte Carlo can perform sensitivity analyses without requiring any additional process simulations. We begin by repeating Eq II.D.2-1 which shows the equation for the reliability with probability distributions:

$$\Re = \int_{\theta \in \mathbf{R}} \Pr\{\theta\} d\theta \qquad \text{Eq II.D.2-1a}$$

$$= \int \ldots \int \Pr\{\theta_1, \ldots, \theta_p\} \; d\theta_1 \ldots d\theta_p \qquad \text{Eq II.D.2-1b}$$

where

$\Re$ is the reliability $(0 \le \Re \le 1)$

$\mathbf{R}$ is the region inside the constraint boundary

The region $\mathbf{R}$ is generally not known beforehand, but this can be worked around by rewriting the equation as:

$$\Re = \int_{-\infty}^{+\infty} 1_{\mathbf{R}}\{\theta\} \cdot \Pr\{\theta\} d\theta \qquad \text{Eq L-1}$$

where

$$\begin{aligned} 1_{\mathbf{R}}\{\theta\} &= 1 \quad \text{if} \quad \theta \in \mathbf{R} \quad (\text{i.e., } \mathbf{h}(\theta) = 0 \text{ and } \mathbf{g}(\theta) \le 0) \\ 1_{\mathbf{R}}\{\theta\} &= 0 \quad \text{if} \quad \theta \notin \mathbf{R} \quad (\text{i.e., } \mathbf{h}(\theta) = 0 \text{ and } \mathbf{g}(\theta) > 0) \end{aligned} \qquad \text{Eq L-2}$$

Eq. L-2 requires process simulations to determine if θ is within $\mathbf{R}$, so it is computationally expensive.

For Conventional Monte Carlo, one generates a number of random points according to the probability distribution, then each point is rigorously simulated to determine whether or not the inequality constraints are satisfied at that location. The fraction of points satisfying the constraints is a statistical estimator of the design reliability, i.e.:

$$\Re \approx \sum_{i=1}^{M} \frac{1_{\mathbf{R}}\{\theta_i^{Sim}\}}{M} \qquad \text{Eq L-3}$$

where

θ_i^{Sim} is the i'th Conventional Monte Carlo random point

M is the total number of Conventional Monte Carlo random points

For a first-order sensitivity analysis, one can write:

$$\Delta\Re = \Re' - \Re = \int_{-\infty}^{+\infty} 1_{\mathbf{R}}\{\theta\} \cdot \left[\Pr'\{\theta\} - \Pr\{\theta\}\right] d\theta$$

$$= \int_{-\infty}^{+\infty} 1_{\mathbf{R}}\{\theta\} \cdot \left[\frac{\Pr'\{\theta\} - \Pr\{\theta\}}{\Pr\{\theta\}}\right] \Pr\{\theta\}\, d\theta \qquad \text{Eq L-4}$$

where

$\Pr'\{\theta\}$ is the new probability distribution

$\Re'$ is the reliability with $\Pr'\{\theta\}$

$\Delta\Re$ is the sensitivity of ← to the parameter distributions

which can be estimated by Conventional Monte Carlo without requiring any additional process simulations:

$$\Delta\Re \approx \sum_{i=1}^{M} \frac{1_{\mathbf{R}}\{\theta_i^{Sim}\}}{M} \left[\frac{\Pr'\{\theta_i^{Sim}\} - \Pr\{\theta_i^{Sim}\}}{\Pr\{\theta_i^{Sim}\}}\right] \qquad \text{Eq L-5}$$

where the same sequence of random points is used as before, for which $1_{\mathbf{R}}\{\theta_i^{Sim}\}$, i=1…m has already been calculated. Fundamentally, one takes the same random numbers previously simulated, but instead of weighting each as zero or one and summing, the nonzero values are reweighted as fractions (the square brackets) and summed.

This calculation should be sufficient for rough first-order approximations. However, it cannot be expected to be accurate for very large differences in the probability distribution, because new regions near the constraint boundary may become significant that have not been sufficiently "blanketed" by the original Conventional Monte Carlo points. For greater accuracy, the analysis may be supplemented by performing a short Conventional Monte Carlo integration with the new probability distribution. Of course, this requires additional process simulations.