REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-09-0144

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
02-01-2009	Final Report	01 April 2006- 30 November 2008

4. TITLE AND SUBTITLE
Research in Evaluation Methods for Data Fusion Capable Tactical Platforms and Distributed Multi-platform Systems in Electronic Warfare and Information Warfare Related Missions

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-06-1-0276

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Llinas, James and Sambhoos, Kedar

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CALSPAN UB RESEARCH CENTER
P.O. BOX 400, 4455 GENESEE STREET
BUFFALO, NY 14225

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force-AFRL,RANDOLPH ST/NA
AFRL, Office of Scientific Research
875 N. Randolph St
Arlington, VA 22203, United States

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distibution A-Approved for Public Release

13. SUPPLEMENTARY NOTES
None

14. ABSTRACT
We have shown that the design of a PE process for any multi-sensor or multi-aircraft fusion based system involves the design of a separate data fusion process involving association and estimation functions for PE purposes per se. Our publications to date have developed the theoretical and architectural groundings for this new PE process, and several case studies have been carried out to show sample implementations of the principles of this new methodology. In addition, some limited objective parametric experiments have also been carried out that show the application of the new evaluation methodology for typical aircraft problems. In this report, we summarize the findings of those past works, and show our research efforts related to extending the design and application of this methodology to air-to air engagement problems involving higher levels of data fusion capability and the employment of electronic warfare systems. The report discusses the detailed strategies for data association, metric estimation and also the analytical techniques that exploit the formality of the methods of Statistical Design of Experiments and Analysis of Variance for these fusion applications. At the end we survey and study the various methods available in literature to solve the large factor Design of Experiment.

15. SUBJECT TERMS
multi-sensor, fusion, theoretical, estimation, aircraft, architectural and warfare

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
			SAR	121	James Llinas
					19b. TELEPHONE NUMBER (Include area code) (617)-645-2357

FINAL TECHNICAL REPORT

Research in Evaluation Methods for Data Fusion-Capable Tactical Platforms and Distributed Multi-platform Systems in Electronic Warfare and Information-Warfare Related Missions

Grant No. FA9550-06-1-0276

Author:
Dr. Kedar Sambhoos
Dr. James Llinas

Program Manager:
Dr. John Schmisseur

Prepared for:
Air Force Office of Scientific Research (AFOSR)
and
Air Force Flight Test Center (AFFTC), Edwards AFB, CA

February 2009

Acknowledgements

This report reflects a research effort that would not have been possible without the support of both AFOSR and AFFTC, and that support is gratefully acknowledged herein, especially the programmatic support of Dr. John Schmisseur of AFOSR and the technical guidance of Mr. Pete Burke, Mr. Dean Baker and Mr. Martin Welch.

Research in Evaluation Methods for Data Fusion-Capable Tactical Platforms and Distributed Multi-platform Systems in Electronic Warfare and Information-Warfare Related Missions

Kedar Sambhoos, James Llinas
Center for Multisource Information Fusion
University at Buffalo,
Buffalo, NY–14260, U.S.A.
sambhoos@cubrc.org, llinas@eng.buffalo.edu

Christopher Bowman
Data Fusion & Neural Networks
1643 Hemlock Way,
Broomfield, CO – 80020, U.S.A.
cbowmanphd@msn.com

Abstract

In previous research, we documented our evolving research that expanded on and formalized an approach to the design of a Performance Evaluation (PE) methodology for Data Fusion (DF)-based tactical aircraft systems. We have shown that the design of a PE process for any multi-sensor or multi-aircraft fusion-based system involves the design of a separate data fusion process involving association and estimation functions for PE purposes per se. Our publications to date have developed the theoretical and architectural groundings for this new PE process, and several case studies have been carried out to show sample implementations of the principles of this new methodology. In addition, some limited-objective parametric experiments have also been carried out that show the application of the new evaluation methodology for typical tactical aircraft problems. In this report, we summarize the findings of these past works, and show our research efforts related to extending the design and application of this methodology to air-to-air engagement problems involving higher-levels of data fusion capability (situation and threat estimation) and the employment of electronic warfare systems. The report discusses the detailed strategies for data association, metrics estimation, and also the analytical techniques that exploit the formality of the methods of Statistical Design of Experiments (DOE) and Analysis of Variance (ANOVA) for these fusion applications. At the end we survey and study the various methods available in literature to solve the large factor Design of Experiment problem, with a detailed guidelines for classification and selection of a proper design.

1 Introduction

Data fusion is an information process involving functional sub-processes that align or normalize the data from several input sources (typically sensor data from surveillance and reconnaissance sensors in defense applications), associate these data to hypothesized specific entities or events or behaviors in the observation space, and then employ these associated or assigned observations toward developing improved state estimates regarding those entities or events or behaviors. Data Fusion (DF henceforth) is a relatively young technology, having had its start in the 1970's driven largely by the need to manage large sensor data volumes from surveillance operations during the Cold War; DF was thus notionalized as a kind of "data compression" technique in the formative years of its development. Later, since many DF applications involved developing these fused state estimates for human users of various type, a more holistic and systemic view of the DF process was developed and the overall process was better defined. The process can be described as follows (Figure 1):

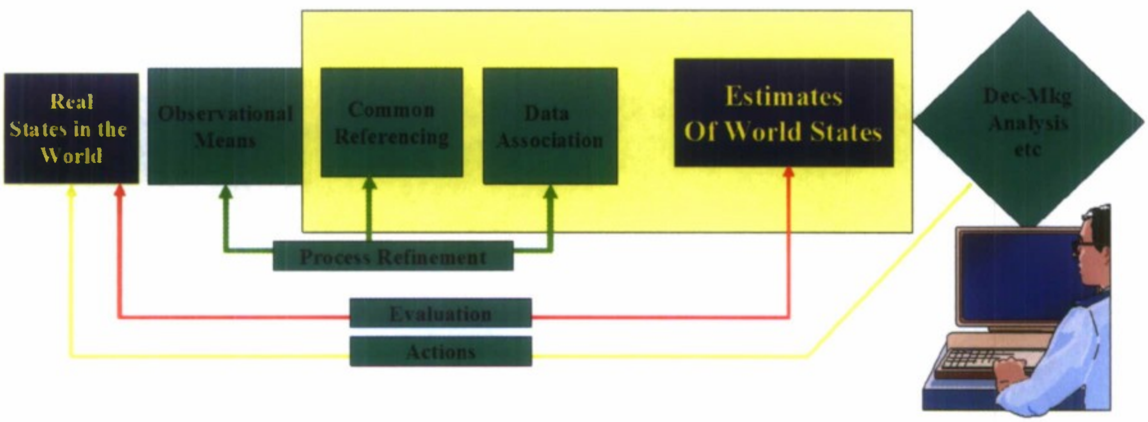


Figure 1. Notional Data Fusion Process.

In the typical defense application context, the state of the Real World is not known but can be estimated a priori, in the usual approach to the design of a deductively-based or model-based approach to the design of an estimation or inferencing process. This dynamic changing Real World is observed by multiple types of sensor systems as noted, and these observations then need to be set into a common frame of reference to include e.g., transforming all observations to a common coordinate system, common time base, etc. Then the observation data set needs to be related to the entities and features in the model knowledge base; this is done via a Data Association process that employs metrics reflecting the "closeness" of any given observation to a modeled entity; this process culminates in the assignment of all observations to the various hypothesized entities. The assigned observations are allocated to specific state estimation algorithms that are specially designed to exploit the overall information content of the assigned observations to generate improved entity state estimates. The usual notion of developing "improved" estimates is in the sense of improvements in accuracy but also in the reduction of uncertainty, as all of the input can sensibly be treated as random variables, resulting from noisy observations from imperfect sensor devices. These estimates, in many but not all applications, are often employed by a human user/operator to effect decisions and action-taking, which themselves can possibly change the Real World state as noted. Ideally, the DF process is designed as an adaptive feedback process, involving for example adaptive, real-time control of the sensors ("sensor management"), or dynamic adaptations to the algorithmic processes ("process refinement") as shown above in the notional feedback depicted. Evaluation of this process is approached within the framework of comparing estimated states of the Real World with the "truth" states of the Real World; truth states are often only known during the Test and Evaluation (T&E) phase of the design and development of the prototype DF process, e.g., during simulation-based or range-testing of the DF process.

A very typical application of the DF process is for the case of multisensor-multitarget tracking, i.e., the case where multiple sensors are employed to develop data that allows DF-based kinematic state estimates (position, velocity, etc) of objects of interest ("targets") to be developed. This is called "Level 1" DF, wherein state estimates on single objects are developed; often these L1 estimates also include identity estimation by fusing the observed entity/feature data to estimate the object class or specific identity type (e.g., fighter aircraft or alternately F-16), depending on the specificity needed in any given application. While these types of estimates are very useful for military applications, as they aid a commander in assessing "where is it?" and "what is it?" more can be done using additional DF methods. The typical next levels of processing, called "Level 2: Situation estimation", and "Level 3: Threat estimation", then leverage from the Level estimates essentially by putting the L1 estimates in context, i.e. exploit contextual information to develop the situational and threat state estimates.

2 Developing an Approach to T&E for Data Fusion Processes

2.1 Addressing the T&E Process for Level 1 Data Fusion-based Tracking Systems

The design and development of algorithmic techniques for estimating the "best" location and related kinematic parameters of moving objects which are observed by single or multiple sensors is a complex process. It is complicated in part by the difficulty of obtaining high-quality measurements from sensor systems due to underlying sensor limitations regarding precision and accuracy, reliability, etc., from natural phenomena that complicate the observing process (weather effects, terrain clutter, etc), and in the defense-problems of interest, from the possible use of sensor countermeasures employed (covertly) by an adversary. Another complicating factor is the inaccuracy associated with the estimation algorithm being used. Virtually all estimation algorithms are model-based, and employ a priori models of target motion, sensor errors, system noises etc in order to estimate the target kinematics. The process is further complicated in environments consisting of multiple closely spaced targets. As a result, there will be differences between the estimated (from the "System Under Test or SUT") and the real ("Truth") picture of the composite multi-object kinematic behavior. The goal of a Multi Target Tracking System (MTTS) designer is to develop a fusion-based tracking system that yields a composite, estimated kinematic picture which is in some sense considered a "good enough" estimate of the composite, true object behavior. Hence at various stages in development of a tracking system it is necessary to evaluate the performance of the system in order to see how close the system's estimate is to the true picture. This is the fundamental issue addressed here: given all the components of a typical tracking system (whose design, as a network of separate fusion processing nodes, is often referred to as a "Data Fusion Tree"), along with the overarching stochastic characteristics of the problem, on what basis can an equitable approach to evaluation of a candidate-design tracking system—the "SUT" -- be based? This issue is far ranging in that it applies to most multi-target, multi-sensor based tracking systems, although strictly speaking it only applies when there is ambiguity in the Data Association process. However, since we lack the knowledge to quantify the degree to which Data Association ambiguity affects the need to carefully determine an evaluation approach, the concern about this issue extends across a broad range of tracking applications.

During the process of designing and developing the SUT DF prototype process, the T&E phases evolve from concept validation testing to developmental testing; in these phases, and often in later in controlled operational testing, the truth states of the "Real World" are known. Presuming the evaluation philosophy is based on comparisons between SUT DF process-generated state estimates and the truth states, the known truth conditions usually allow for straightforward calculation of the various evaluation metrics being employed. Such evaluation techniques presume that there is an ability to relate specific SUT-generated state estimates to specific truth states, i.e., that the associability between "Tracks" (the SUT-generated track estimates) and "Truth" (the specified Truth states for the given T&E experiments) is known. However, conditional on many factors both related to the sensors being employed as well as the behavior of the targets and also the specific characteristics of the various algorithms being employed, the ability to clearly determine which SUT Track should be compared, for evaluation purposes, with which Truth track may often be unclear. Such ambiguities in DF-based tracker algorithm evaluation have been known and flagged as evaluation issues as far back as the 90's (e.g., see Refs [1], [2]). However, very few papers describing techniques to deal with this problem when evaluating fusion-based trackers have been published, and in particular almost no papers (other than our past works) have been published that address the T&E methodological issues related to this problem.

There are a number of types of tracker algorithm pathologies that can arise can give rise to the Track-to-Truth association problem. A few cases are shown below in Figure 2.

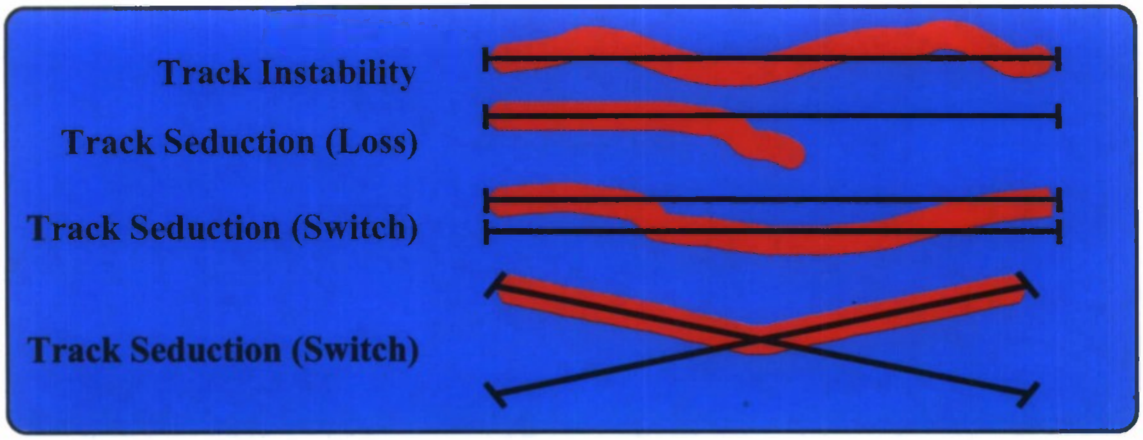


Figure 2. Examples of pathologies in tracker estimates.

Most of these problems occur as a result of mis-associations of the sensor data to the estimates being generated by the tracker algorithm (often based on Kalman Filter-based techniques), such that some of the sensor data for given targets are associated to another target, or because a local association or estimation error creates a condition where the track is lost, or another example is when two (or "n") targets are truly closely-spaced and sensor resolution limits coupled with association errors result in track switching, where the estimation process mixes estimates for multiple targets together (this can occur even when target identity is also being estimated although such estimates do help in reducing this particular type of error). A wide variety of other difficulties can arise even with the most sophisticated and modern tracking techniques. Thus, tracker evaluation conditions such as shown in Figure 3 can arise:

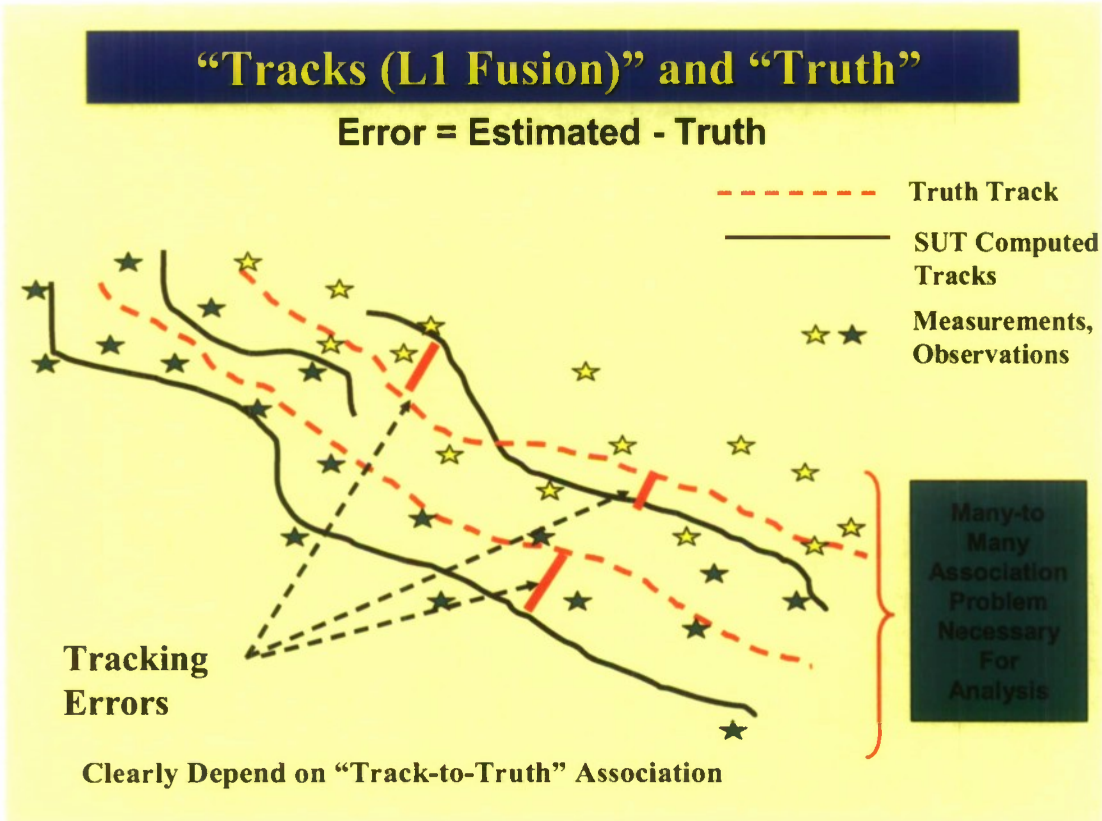


Figure 3. Notional evaluation case for fusion-based target tracking process.

The point of this diagram is that in order to assert that specific tracking errors exist, i.e., that there are specific differences between the SUT DF-based track estimates (here shown in black) and the Truth tracks (in red), an assertion of which SUT track goes with which Truth track must be made. In the face of the many pathological conditions that can arise (e.g., the track fragment in Figure 3, as well as the closeness of the computed tracks to the Truth tracks), such associations are not at all easy to assert. As the figure indicates on the right-hand side, a many-to-many association problem must be solved to assert the SUT Track to Truth track relationships with any confidence. The specific insertion of such steps is one specific recommendation of our proposed T&E methodology. Said otherwise, a new Data Fusion process must be designed for the specific purpose of testing and evaluating any DF-based tracker. (We will show later that this is a requirement for any DF process, including the higher levels of fusion (L2, L3) as previously described.) This modified approach to T&E is shown notionally in Figure 4.

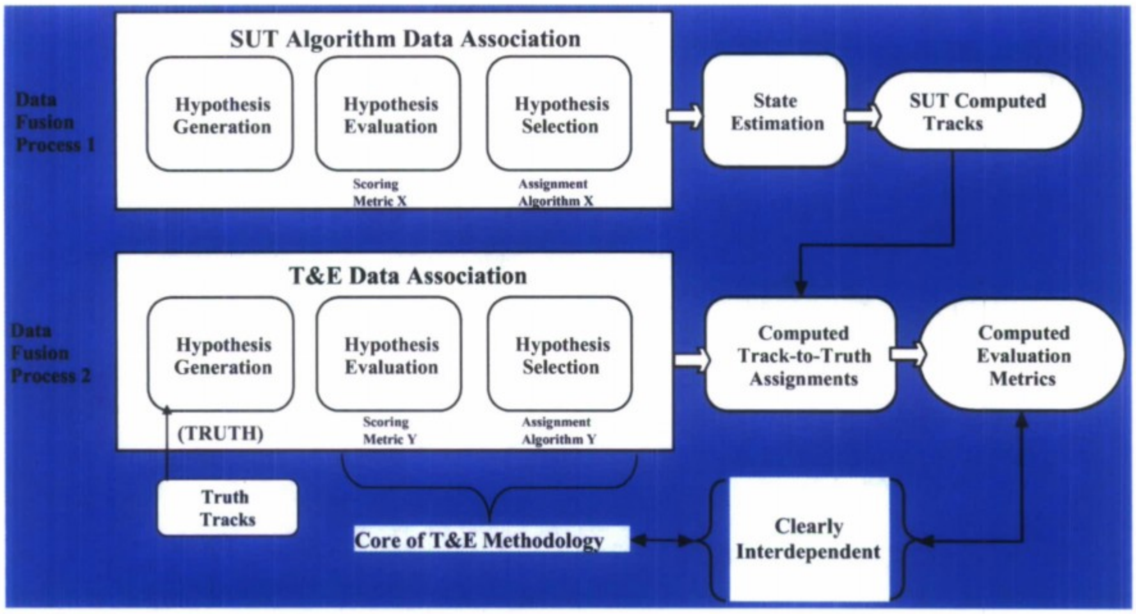


Figure 4. Separate Data Fusion processes; the SUT DF process and the T&E DF process.

It should first be mentioned that the approach to constructing any Data Association (DA) approach involves three sub-processes as shown in Figure 4: Hypothesis Generation, where the feasible causes of any given observation, to include Electronic Warfare techniques i.e., deception, are defined, Hypothesis Evaluation, where metrics or scores that reflect the degree of "closeness" of an observation to an estimate are defined and calculated, and Hypothesis Selection, where the many-to-many DA or "assignment" type problem is solved, culminating in an optimal assignment of the observations to the appropriate state estimation (fusion) algorithms for each target. This last step employs what are called "assignment algorithms" imported from the field of Operations Research. (The term "hypothesis" here means and association hypothesis, i.e., a nominated observation-to-estimate pairing.)

Along the top of Figure 4 we have the DA and DF process for the System Under Test (SUT); this process notionally uses certain Scoring Metrics and Assignment algorithms. This process operates on the multi-sensor input stream and produces target track estimates, the SUT Tracks. The T&E DA process, notionally employing different types of Scoring Metrics and Assignment algorithms, takes the Truth tracks as the definitive associable hypotheses, and calculates the "best" assignments of SUT Tracks to

Truth Tracks. Given those assignments, the tracking errors (basically grounded on differences between the estimated states and the Truth states) and any nominated performance metrics can be estimated. As pointed out in Figure 4, the values of the performance metrics clearly depend on the computed Track-to-Truth association.

In all of the above, we have been emphasizing that the DF process produces estimates. This is because, in the strictest sense, the inputs to the DF processes are the statistically-noisy sensor data having stochastic properties. These features have yet other implications for the T&E methodology, namely that the stochastic nature of the process needs to be recognized and dealt with in any T&E approach. At least when conducting any simulation-based T&E, this implies that (a) the experiments should be designed through the employment of the methods of statistical experimental design (aka Design of Experiments or "DOE"), and (b) in conjunction with this that Monte-Carlo based replications of any given test condition should be done.

It is recognized by the way that such rigor comes at a price, even when using simulations, and especially when doing field tests and the like. It is likely that there has been limited application of these formal methods because of the cost implications. As academics however, we feel it is our job to nominate rigorous methods so that their application can at least be assessed in any given case. It is only through the use of such methods that assertions about the computed metrics can be made with statistical confidence.

There are other issues regarding the design of the overall T&E process, and a complete discussion of them is beyond the scope of this paper. To give one example, there is the issue of the design of the overall T&E process for a typical prototype SUT DF process. Any given real DF process will involve a complex processing architecture, not a single DF node, because the design of such real systems involves various design tradeoff decisions. A typical DF process architecture may appear as that in Figure 5.

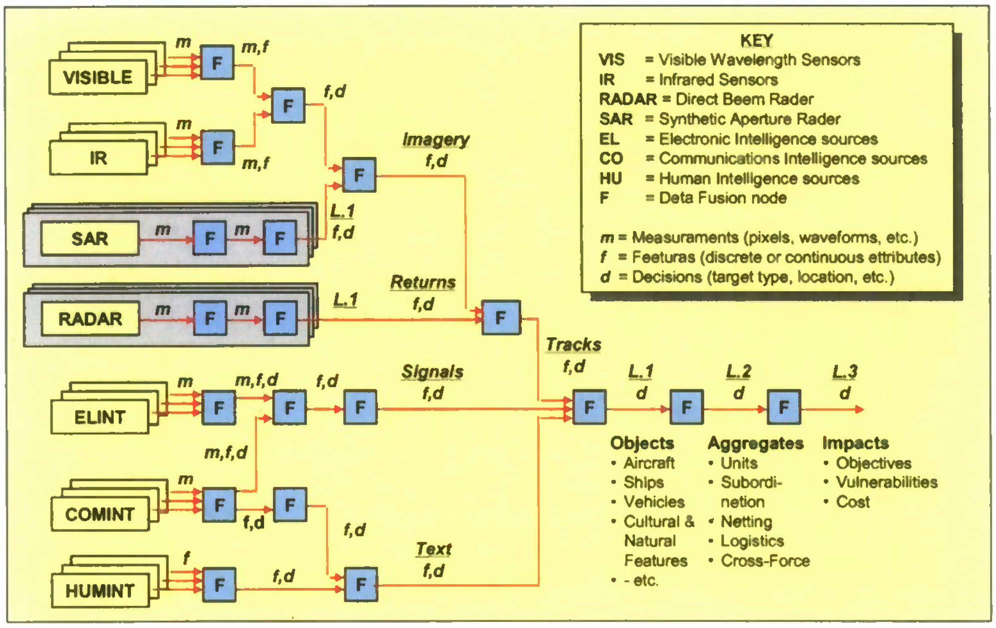


Figure 5. Representative Data Fusion processing architecture.

Here we see that the processing flow for a multisensor system involves the usual batching and partitioning decisions necessary to evolve an effective and efficient processing approach. It is thus typical that tuples of sensor groups are joined in local DF operations (each "F" node above is a fusion operation); this is done for various reasons, to include availability of the data, or commonality of the data, etc. The point is that there are in any system multiple DF nodes and so the strategy for the design of the T&E DF process can become equally complex. Examples of the various strategies that can be applied to the design of the T&E DF process are shown in Figure 6.

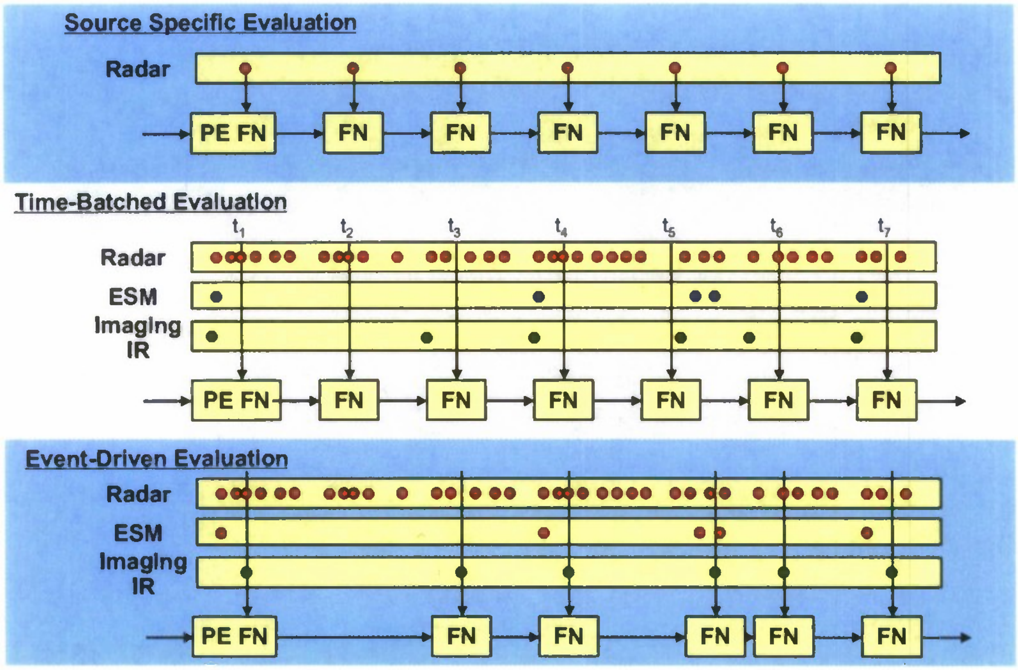


Figure 6. Representative design choices for T&E DF nodal processing.

Here we see that the T&E-specific DF processing architecture may involve choosing DF nodes that are sensor or source specific (e.g., specific to a radar sensor), or a strategy that is time-based where Track-to-Truth associations and estimates of metrics are computed at set time intervals, or a strategy that is event-driven according to the events occurring in the test scenario. Thus, the assertions regarding the Track-to-Truth relations can change over time, or change according to the flow of events, or according to yet other choices in the overall T&E DF processing architecture.

2.2 Addressing the T&E Process for Higher Level Data Fusion-based Tracking Systems

In this research, we are now looking at the higher levels of fusion, involving the formation of Threat or Risk estimates for each friendly aircraft in these scenarios, as developed from the available multi-sensor data. In this case we are exploiting the use of the onboard radars and the Electronic Support Measures (ESM) sensors that estimate the operating modes of hostile radars. Conceptually, the Actual Risk to a friendly aircraft can be thought of as defined by the relationship between an Inherent Risk and the ability to thwart that risk with available countermeasures. The way this notional process is being implemented is shown in Figure 7.

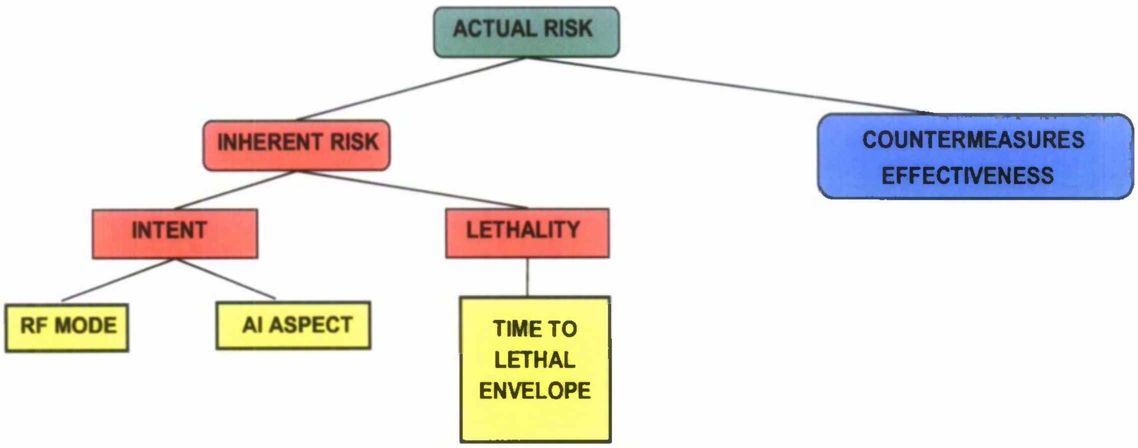


Figure 7. Design of the Actual Risk estimation logic.

For the cases we are examining, the Inherent Risk can be estimated by estimating the Intent and Lethality of a given hostile platform. In turn, with the available sensor suite, we can employ that data to estimate Intent by examining (a) the Mode of the hostile radar (from the ESM data) and (b) the relative or Airborne Intercept aspect or inter-platform geometry, to assess for example whether the hostile is in a shoot geometry; this can be estimated from the estimated track data for the hostile and the ownship navigation data. We estimate Lethality using a concept called "Time to Lethal Envelope" or TTLE, also estimable from the fused kinematic data, and representing a hypothetical worst-case condition where both platforms turn directly to each other, with the TTLE being the time it takes to get within maximum hostile weapon range.

The design of Airborne Intercept aspect or inter-platform geometry is shown in Figure 9. The Airborne Intercept is calculated based on the truth track of the own ship (blue). The aspect angle is the angle formed by the estimated range vector with the estimated velocity vector of the hostile platform (red). The Airborne Intercept is the angle formed by the own ship velocity vector and the hostile ship velocity vector. (Shown as Tue AI Aspect Angle C and Estimated AI Aspect Angle D in Figure 8)

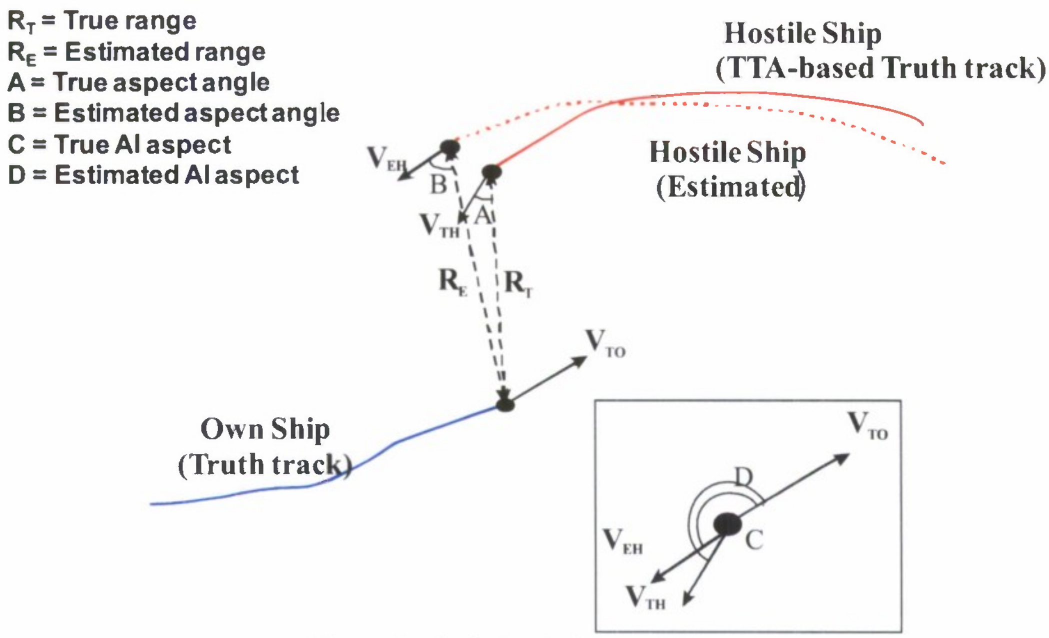


Figure 8. Air Intercept Aspect Geometry

To calculate the Intent the possible scenarios should be considered. These scenarios are represented in Figure 9. Here the range vector is defined to origin at the estimated hostile location and end at own ship location. The range vector angle is calculated based on these two locations. There are four possible scenarios; based on the AI Aspect angle formed: less than 90^0; between 90^0 and 180^0; between 180^0 and 270^0 and between 270^0 and 360^0. There are two possibilities in each of the case and the range angle is used to differentiate between a hostile and friendly situation. These possibilities are shown in Figure 9.

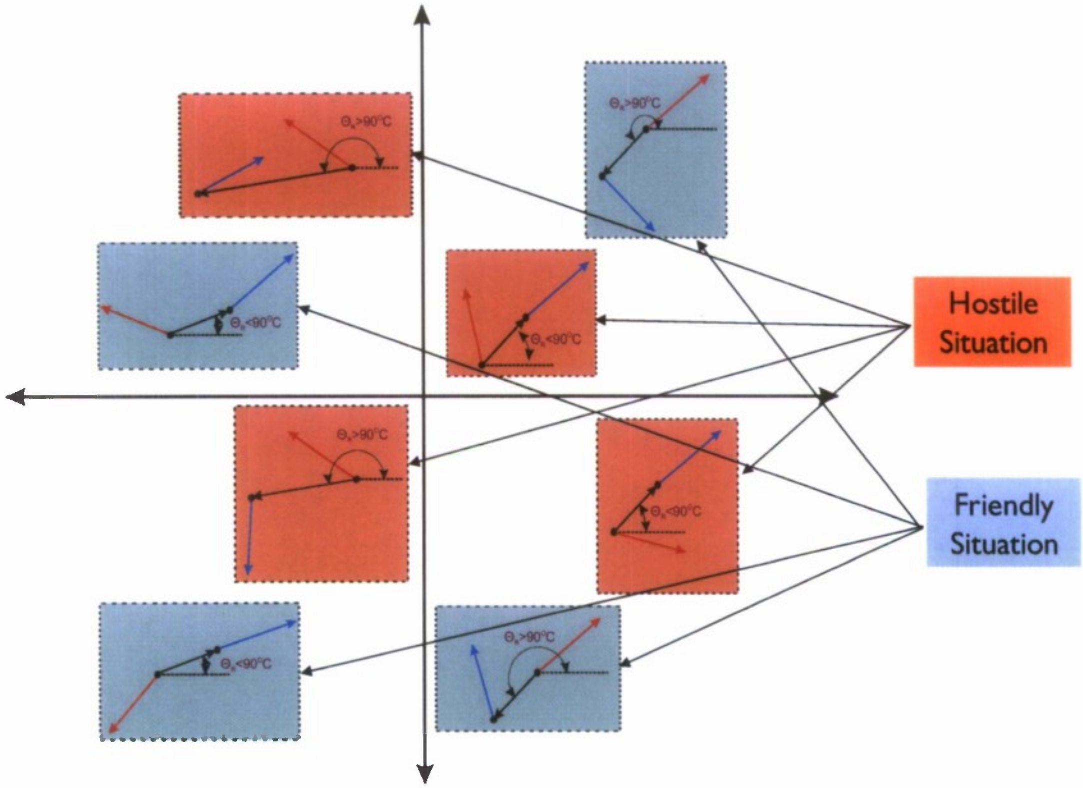


Figure 9. Air Intercept Aspect Geometry – Possible scenarios

To further simulate hostile concept of employment or operation (COE/COP) for hostile radar, we use the ESM data. The ESM data is estimated using the estimated velocity vector and field of view (FOV) [FOV angle is a user input] of the hostile platform (30^0 around the velocity vector). In Figure 10 we see that, hostile H can see platform F_1 while F_2 is out of the FOV range of H.

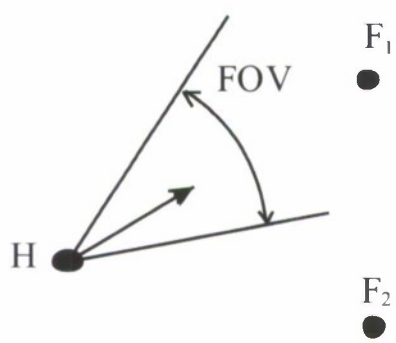


Figure 10. Ownship/ Friendly ESM

There are four possible modes simulated for the hostile radar:

1. Unknown: F is not in FOV of H

2. Search: When F is in H FOV and both platforms are in an approaching or closing kinematic relationship.
3. Track: When F is in search mode for Δt_1 time.
4. Lock-on: (Figure 11) When F is in Track on mode for Δt_2 time and also reaches within the Hostile missile range (R_{max})

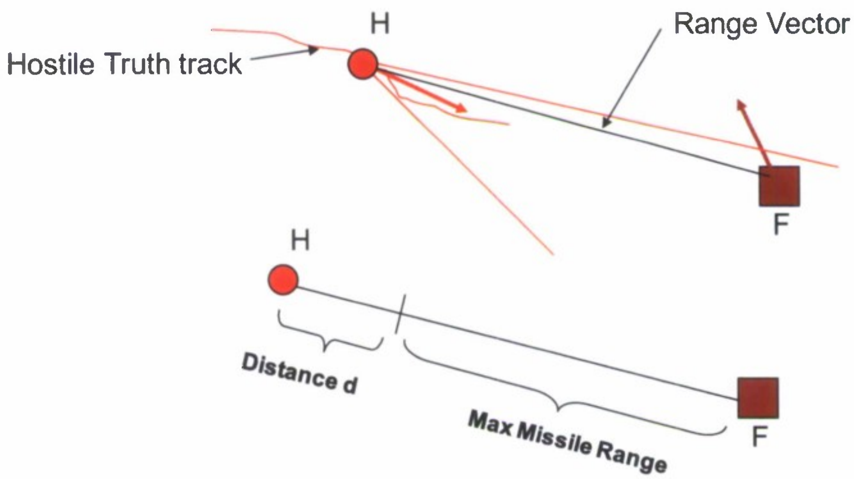


Figure 11. Lock-on Mode Declaration Logic

The ESM sensor is assumed to have long range, 4π sensitivity. Being passive, the correct calculation for the condition that the Friendly has ESM data available to it is based on whether the Friendly is truly in the FOV of the Hostile radar (we assume the H ConOp employs the radar actively). Thus, the geometric calculations for FOV containment are based on the Hostile Truth track data. To generate True Mode according to this logic, go to Confusion matrix to generate Actual Mode Declaration Report (Table 1).

Table 1 . Actual Mode Declaration Report.

Actual Mode Declaration	True Mode (Geometry, Time)		
	Search	Track	Lock-On
Search	P(S|S)	P(S|T)	P(S|LO)
Track	P(T|S)	P(T|T)	P(T|LO)

Lock-On	P(LO|S)	P(LO|T)	P(LO|LO)
Unknown	P(U|S)	P(U|T)	P(U|LO)

To simulate the lethality, we use the concept of "Time to Lethal Envelope" (TTLE). TTLE is the hypothetical, worst-case calculation which assumes that both hostile and friendly platforms turn directly toward each other at max velocity. TTLE is the time to close to within Hostile missile launch maximum range depicted in Figure 12.

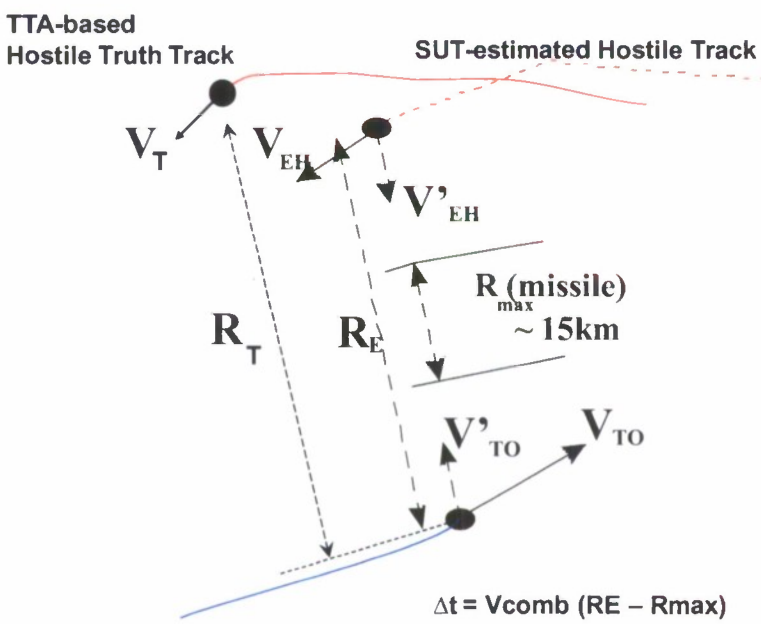


Figure 12. Time to Lethal Envelope (TTLE)

These Intent and Lethality estimates can be logically fused to assert a level of Inherent Risk. The top-level approach is shown in Figure 13. Here we see that the friendly platform sensor data are fed to the Level 1 SUT DF process which provides the fused kinematic state estimates on all hostile platforms. The ESM data are used to estimate the RF Mode of the hostile radars and the fused kinematic estimates are used to construct the Level 2 SUT DF estimate of the "situation", and then the situational estimates are fused using the logic of Fig.9 to provide the Level 3 fused estimate of Inherent Risk. We are using a Fuzzy Logic approach to forming the L3 DF logic for the reasons shown in Figure 13.

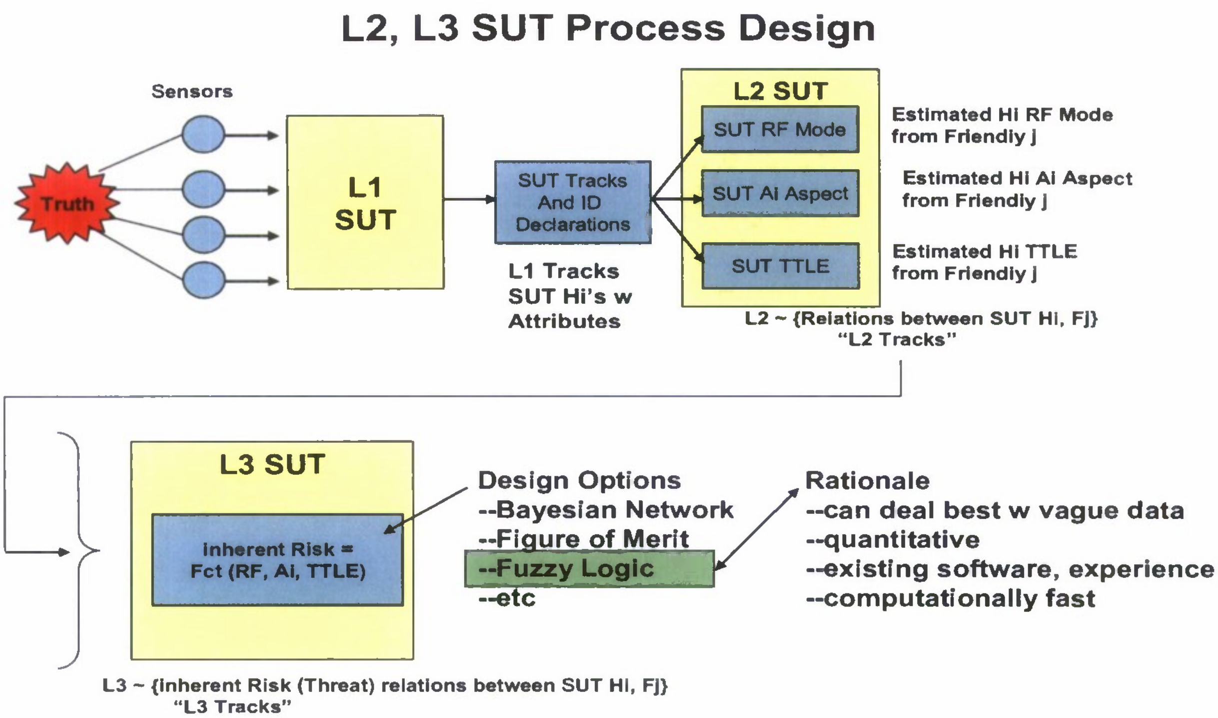


Figure 13. Top-level approach to higher-level fusion process design.

There are various design options available to fuse the L2 SUT estimates to get L3 SUT estimate of Inherent Risk; like Bayesian Networks, Figure of Merit, etc. Here we use Fuzzy Logic to fuse the L2 estimates. Fuzzy Logic (FL) is an inferencing methodology that is directed toward vague relationships between evidence and assertions. Using natural language statements that contain appropriately-vague terms (e.g., "close"), FL provides a quantitative framework for relating the interdependent phrases of these expressions. Fuzzy inference is the process of formulating the mapping from a given input to an output using FL. Fuzzy inference systems that have been successfully applied in fields such as automatic control, data classification, decision analysis, expert systems, and computer vision. Because of its multidisciplinary nature, fuzzy inference systems are associated with a number of names, such as fuzzyrule-based systems, fuzzy expert systems, fuzzy modeling, fuzzy associative memory, fuzzy logic controllers, and simply (and ambiguously) fuzzy systems.

Fuzzy Logic is a useful tool in making decisions in light of information that is imprecise and incomplete. Given information collected from multiple sources such as a target's location, aspect, and speed, FL can be used to measure the degree of danger of the target without formulating complex mathematical equation. The FL functions are more natural for the representation of the feeling of incertitude. A very precise information is not expected absolutely, but hope for the greatest possible coherence. On the other hand, precise but fluctuating data more usually result from the observation of a physical phenomenon.

Given information collected from multiple sources such as a target's ESM Mode, Air Intercept, and TTLE, we measure the degree of Inherent Risk of the target by adopting FL easily without formulating the complex mathematical equation. In this research, we adopt FL for measuring the Inherent Risk because of a couple of reasons. In certain observation and reporting circumstances, it may not be appropriate to represent those variables in the probability domain. That is, the state of these variables does not have an associated "crisp" set. For convenience, we suppose that the three SUT estimates have three states respectively; these are the type of representations that would come from a reporting or message-based input, rather than from the sensors themselves. Unlike the variables in Bayesian Networks, which have crisp set, these Aspect and TTLE variables have fuzzy sets.

Consider an example of measuring the height of all the children in a class. How can we classify the height of the children as low, medium or high? If we say a person having height greater than 6 feet is tall, then we have a situation where a 5.9 feet person terms out to be of medium height and a person with height of 6.1 feet turns out to be in tall group. But in FL a *membership function* is used to measure the degree of membership of the quantitative value (Here, height of a child) in a fuzzy set.

Fuzzy Logic also has advantage in representing of kinematic and angle data since it uses natural language. The fuzzy inference system has membership functions, fuzzy logic operators and if-then rules.

There are two types of fuzzy inference systems: Mamdani-type and Sugeno-type. The Mamdani type inference system has been adopted in the framework for L3 fusion that is reported herein.

Table 2. Fuzzy Logic - Threat/Risk Logic Rules

	AI	TTLE	RF	Threat		AI	TTLE	RF	Threat
1	low	large	Unknown or search	low	15	med	low	Track	high
2	low	med	Unknown or search	low	16	med	large	Lock-on	med
3	low	low	Unknown or search	med	17	med	med	Lock	high
4	low	large	Track	low	18	med	low	Lock	high
5	low	med	Track	med	19	high	large	Unknown or search	low
6	low	low	Track	hi	20	high	med	Unknown or search	low
7	low	large	Lock-on	med	21	high	low	Unknown or search	med
8	low	med	Lock	high	22	high	large	Track	med
9	low	low	Lock	high	23	high	med	Track	high
10	med	large	Unknown or search	low	24	high	low	Track	high
11	med	med	Unknown or search	low	25	high	large	Lock-on	med
12	med	low	Unknown or search	med	26	high	med	Lock	high

13	med	large	Track	med	27	high	low	Lock	high
14	med	med	Track	high					

Note that the L3 SUT DF process can be thought of as producing "Inherent Risk Tracks", i.e. timewise histories of the estimated level of Inherent Risk, conceptually as much as track as the timewise position histories of a physical platform. Note too that the issue of associating estimated Inherent Risk tracks-to-Inherent Risk Truth tracks will be a challenge in designing the new T&E DF process to evaluate these new estimates. The Inherent Risk Truth tracks are those computed by using the truth values of kinematics etc in the track formation.

2.3 Addressing the T&E Process for Level 4 Data Fusion-based Tracking Systems

That risk would be mitigated according to the possible employment of Countermeasures (CM) available to the friendly platform. Continuing from the T&E framework for Level 2 and 3 Fusion process we employ the Level 4 Fusion process as shown in Figure 14. Here we employ electronic countermeasure (ECM). Any electronic effort which intends to disturb normal radar operation is referred to as ECM. ECM are employed to accomplish improper or delayed target detection, analyst deception or generate false positives. There are two classes to CMs, one which actively deny radars to perceive an measurement like Jammers, and deceptive CMs like changing trajectory.

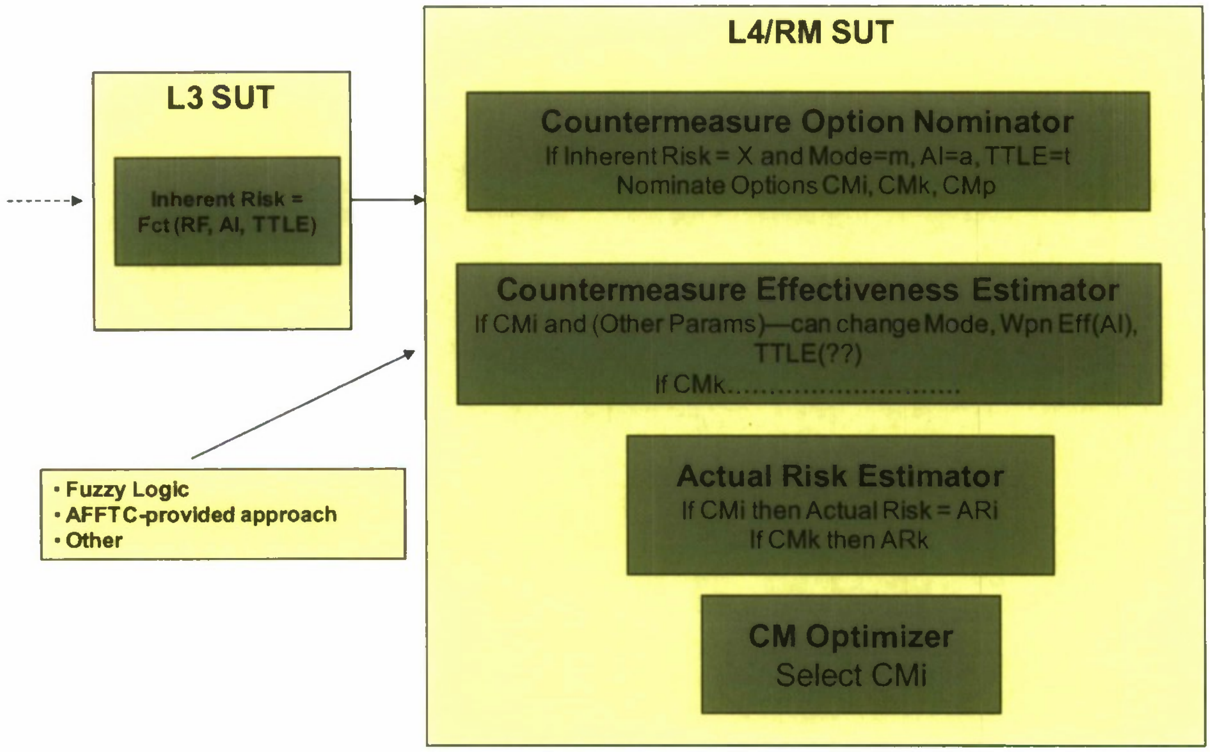


Figure 14. Top-level approach to level 4 fusion process design

Here we simulate the situation where the own ship platforms have Self-Screening Jammers (SSJ). They are type of self protecting jammers which carry a jammer on the platform like beam jammers. Another type of jammer not implemented here is Stand-off jammer which requires an escort vehicle which carries the jammer. The details about the jammers parameters were referenced from Mahafza and Elsherbeni [49]. Here are some of the terms relevant to the CM implementation [49]:

P_t = Peak Power in Watts = = 50.0e+3

G = Antenna Gain in dB = 35.0

λ = Antenna wavelength

A_r = Antenna aperture

σ = 10.0;   % Radar Cross Section in m sqaured

B_r = Radar Operating Bandwidth in Hz = 667.0e+3

L = Radar Losses in dB = 0.1000

P_J = Jammer Peak Power in Watts = 200.0

B_J = Jammer Operating Bandnwidth in Hz = 50.0e+6

G_J = Jammer Antenna Gain in dB = 10.0

L_J = Jammer Losses in dB = 0.10

R = Range

A single pulse power received by the radar from target is given as:

$$S = \frac{P_t G^2 \lambda^2 \rho \tau}{(4\pi)^3 R^4 L}.$$

The power received by the radar from a SSJ jammer at same range is given as:

$$J = \frac{P_J G_J A_r}{4\pi R^2 B_J L_J}$$

Where $A_r = \frac{\lambda^2 G}{4\pi}$ then $J = \frac{P_J G_J \lambda^2 G}{(4\pi R)^2 B_J L_J}$. A jammer can be identified by its effective operating bandwidth and its Effective Radiated Power (ERP):$ERP = \frac{P_J G_J}{L_J}$. Then the Signal to Jammer ratio(S/J) is given as:

$$\frac{S}{J} = \frac{P_t * \left(10^{\frac{G}{10}}\right) * \sigma * B_J * \left(10^{\frac{lossj}{10}}\right)}{4.0 * \pi * P_J * \left(10^{\frac{G_J}{10}}\right) * B_r * \left(10^{\frac{L}{10}}\right) * ((R * 1000.0)^2)}$$

S/J tells what the effect of the jammer on the radar is. The jamming power from the platform is a one-way transmission while the target signal power is a two way transmission. So, the jamming power is generally greater than target signal power (S/J < 1). But as the target comes closer there will certain range at which S/J ratio equals 1, and it is called as cross-over range and is given as:

$$Cross\ Over\ Range\ = \frac{\sqrt{\left(\dfrac{P_t * \left(10^{\frac{G}{10}}\right) * \sigma * B_J * \left(10^{\frac{L_J}{10}}\right)}{4.0 * \pi * P_J * \left(10^{\frac{G_J}{10}}\right) * B_r * \left(10^{\frac{L}{10}}\right)}\right)}}{1000.0}$$

So, to remain undetected the platform has to remain at a range higher than cross-over range. The range at which the radar can detect and perform proper measurements for a given S/J is called as burn-through range. So the logic on CM implemented here is that the own ships try to maintain safe distance from the hostile targets. The platform jamming power is in proportion to the target signal power until the platform reaches the cross-over range. When the platform reaches the cross-over range and is unable to block the target signal, the second counter measure logic sets in. The ownship platform will change its flight plan by changing the flying trajectory by a preset angle and try to move out of the hostile field of view (FOV).

3 Case Study: PE Simulator for AFFTC

In our earlier work, we have summarized some of the generalized issues when considering the test and evaluation of a prototype data fusion process (what we have called the "System Under Test" or SUT). We focused on the problem of PE and the "fairness" issue for the distributed data fusion case. In this AFFTC application, one type of distributed fusion application will be the case of multiple aircraft platforms working cooperatively on a common mission, each performing local or platform-specific fusion while also exchanging data and fused estimates to each other. A core evaluation issue herein is the assessment of the degree of consistency in the multiple track pictures across the platforms. It is also critical to guarantee that the alternative PE network node outputs are consistent, in accordance with a consistency specification pertinent to the application. PE nodes perform track-to-truth association to support track accuracy-related or other MoPs, and perform track-to-track association to support platform track file consistency-related MoP estimation for two or more internetted platforms (e.g., Joint Strike Fighters (JSFs), or the "F-35" aircraft).

3.1 PE Node Design

In the PE framework the PE nodes perform 3 necessary functions: (i) data preparation (ii) data association and (iii) MoP state estimation. In our Case Study, during data preparation the PE node puts tracks and truth information in [x, y] co-ordinates and common time. Data association performs deterministic track-to-truth association and track-to-track association. During data association the following three actions are performed:

(i) Hypothesis Generation,

(ii) Hypothesis Evaluation, and

(iii) Hypothesis Selection.

The PE node uses a Kalman filter for Level 1 MoP state estimation. Using the Level 1 estimates the Level 2 and 3 estimates are generated. SUT tracking errors will induce Intent (AI Aspect and ESM mode) and Lethality (TTLE) errors.These errors will propagate to a component error in Threat/Risk assessment. Note that these are different than Threat errors derived from Threat-to-Truth Threat Association.

3.2 Case Study Measures of Performance for PE

This overall T&E methodology has been applied to some cases of interest to the Air Force that involve DF-capable tactical aircraft. (Many modern-day tactical aircraft have multiple sensor systems and can employ DF processes to support the execution of their missions.) Because certain aircraft employment concepts involve multiple aircraft cooperating on a given mission, we have recently examined a case involving two friendly aircraft engaged in an air-to-air combat environment with six hostile aircraft. From a Data Fusion point of view, this involved addressing the T&E issues associated with Distributed Data Fusion (DDF), since fusion occurs not only on each friendly aircraft but also between them, since they are intercommunicating, and exchanging both data and fused estimates in a rather complex DDF process. The framework of these experiments is shown in Figure 15.

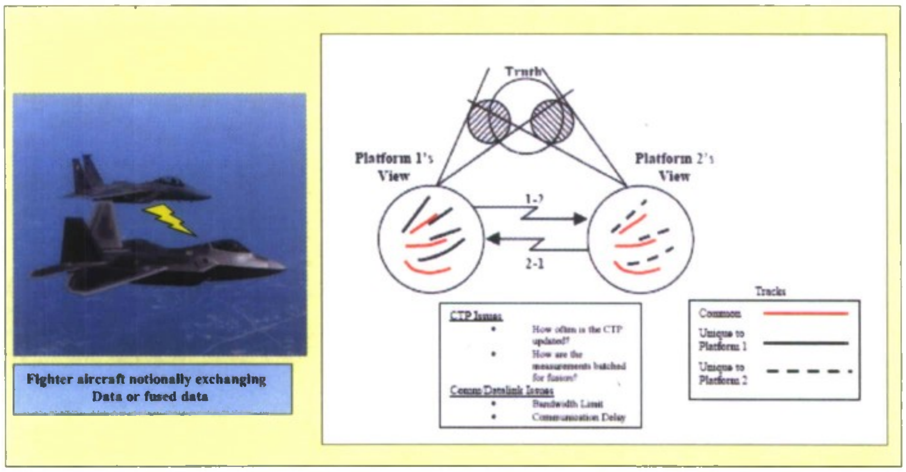


Figure 15. Framework of Distributed Data Fusion experiments.

From the point of view of supporting the tactical mission, one critical issue of course if whether there is a consistent "track picture" across the two aircraft. It can be seen in Figure 15 that it is typical that there are differences in the local target track pictures on each platform which need to be reconciled for mission application. In these recent studies then, we studied fused track picture consistency as a function of certain factors, looking at both Track-to-Truth and Track-to-Track consistency metrics.

Figure 15 depicts how the two platforms have their own view of the truth picture based on the on-board sensors. There are both "common" pictures and "unique" pictures. Let us assume, for the sake of example, that all the on-board sensors see the same targets. Let platform 1 sees 3 tracks (based on on-board sensors) which are common to platform 2 and vice versa. The common tracks are shown in red. Note that even though both of the platforms see the same targets, their measurements about those common targets could be different depending on how the on-board sensors report the measurements. Also there are certain targets that are uniquely seen by platform 1 and platform 2; note that some of either the common or unique tracks could be false tracks.

Each of the platforms exchange their track files and data fusion is done upon receipt of this information at each platform. We will explain further how this information is exchanged when we discuss Tier 0, Tier 1 and Tier 2 (Section 3.3). We assume that there is no bandwidth limitation in communication. We have incorporated more realistic asynchronous (delayed) communication among the sensors and the platforms.

The baseline distributed fusion output is the Consistent Tactical Picture (CTP). The sensor track file "consistency" is computed at each time point as the percentage of matching CTP tracks in the track files of each platform. In addition to this measure, the following four higher level consistency metrics have been computed:

1. *Track-to-Track Aspect, ESM and TTLE Consistency*: These are Level II metrics which compares the Air Intercept, ESM mode and TTLE estimates across platforms.
2. *Track-to-Truth Aspect, ESM and TTLE Consistency*: These are Level II metrics which measure the accuracy of the Air Intercept, ESM mode and TTLE estimates.
3. *Track-to-Track Threat Consistency*: This is a Level III metric which compares the estiamted Threat across platforms.
4. *Track-to-Truth Threat Consistency*: This is a Level III metric which measures the accuracy of the estimated Threat.

3.3 Explanation of Fusion Tiers

Tier 0: (Figure 16) In Tier 0, each of the on-board sensors (Radar, ESM and IRST) fuse their own reports. The resultant Tier 0 tracks are then fused together to get the Tier 1 consistent track picture. Here the information is not yet shared across the platforms, so the result tends to be less accurate than for example the fusion of Tier 0 sensor tracks to the all source CYP. Generally, batching of larger data sets for fusion is more accurate; albeit more complex.

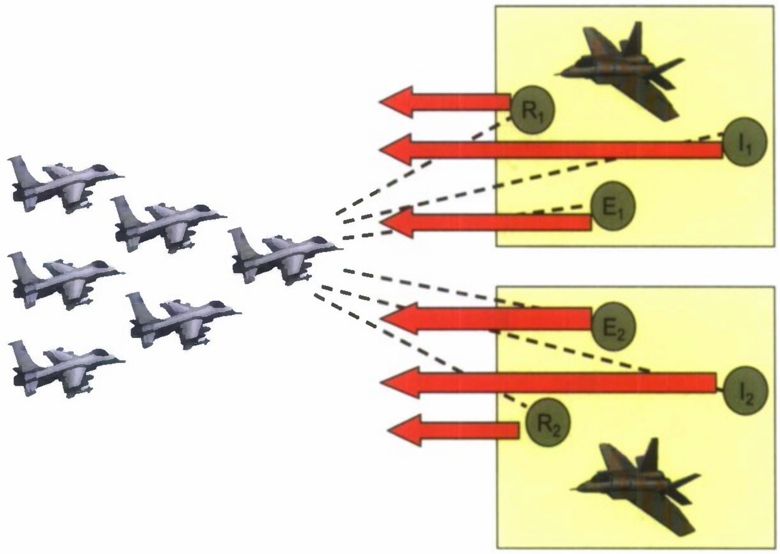


Figure 16. Tier 0

Tier 1: (Figure 17) In Tier 1, each of the on-board sensors (Radar, ESM and IRST) share their Tier 0 track files to generate the ownship consistent track picture. This is typically done for each sensor track file as it is updated, rather than all sensors at once. The DNN architecture exposes these and many other ways to network fusion nodes on a single platform for Tier 1 fusion or on multiple platforms for Tier 2 fusion.

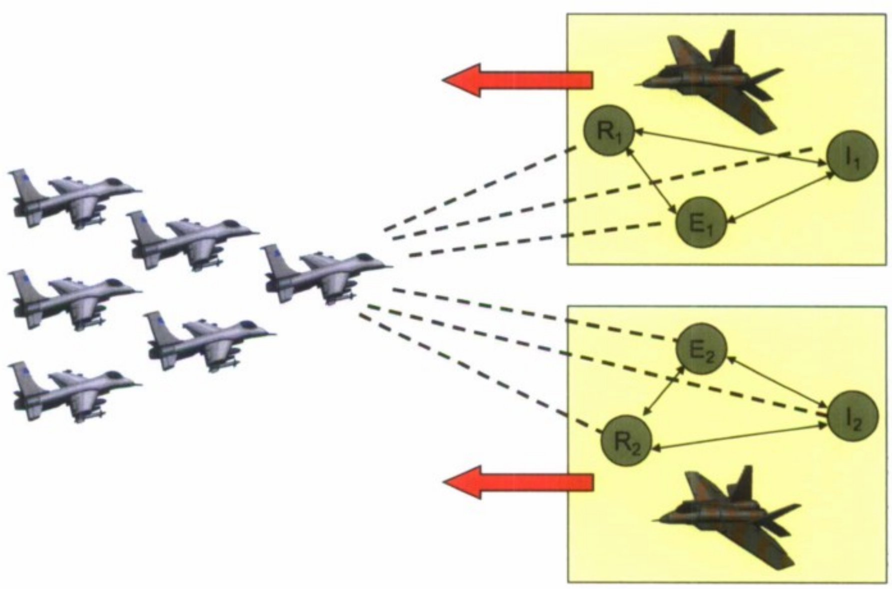


Figure 17. Tier 1

Tier 2: (Figure 18) In a typical Tier 2 fusion the Tier 1 track files are fused sequentially as each Tier 1 track file is updated. A modified form of a Tier 2 fusion network is for each platform to share its own sensor measurements with the other platforms. This can be done one sensor at a time sequentially as each sensor scan of data is received. This alternative tends to be more accurate, however at a cost of more communications bandwidth and fusion complexity (e.g., due to report propagations for time delays, multiple platform coordinate misalignments, internetted ghost tracks, etc.).

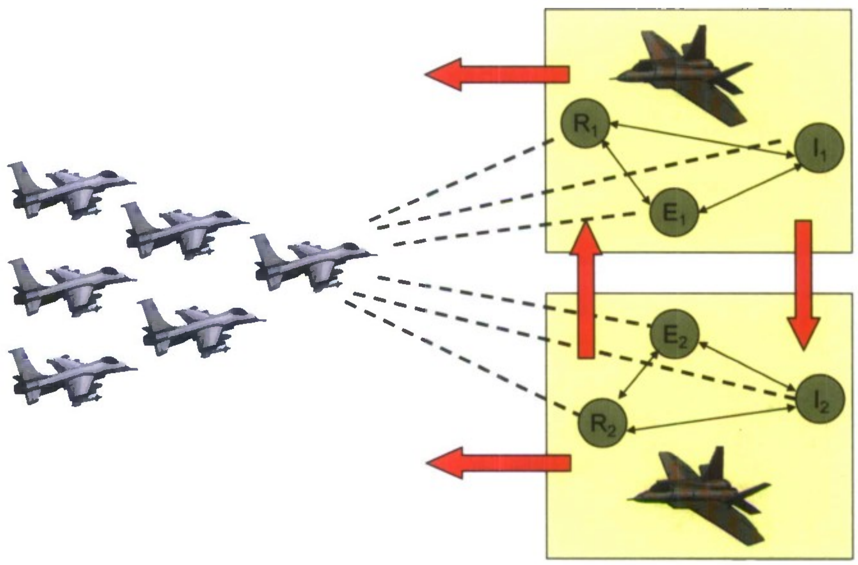


Figure 18. Tier 2

4 Experimental Results and Analysis

The baseline 2 vs. 6 offensive sweep scenario has 6 foe fighters (targets) engaging simultaneously in pairs from left and right 45 degrees and center to achieve a simultaneous missile launch against the blues (platforms). The blue 1 launches AMRAAM missiles on reds (fighters) 1, 2 and 3, 4 respectively. The blue 1 launches AMRAAM for the second time against the surviving red. Then the other blue turns towards reds 5, 6 and launches missile. All the red fighters are in a pair staggered formation with the trailing fighter off to the left or to the right, sufficient to be not resolvable by blue radar until after the final red turn.

Tier 0: We ran the simulation for all Tiers from time periods 1 to 329 with an interval of 1 time period. The time period was 1 second. The baseline 2vs6 offensive sweep scenario has 6 foe fighters coming towards 2 blue fighters with the objective of engaging at 10-15 km simultaneously in pairs from ±45 degrees and center. The blue launch AMRAAM missiles between 20-25 km on 1, 2 and 3, 4, respectively.

The second launch by blue 1 against the surviving red 3 occurs at about 10-15 km. Then the other blue turns at 5g towards red 5, 6 and launches on 5, 6. All fighters are in a pair staggered formation with the trailing fighter off to right and behind sufficient to be not resolvable by blue radar until after the final red turn.

The blue and red fighters are both initially in search mode for each other. Once the reds detect they turn off emissions and execute their pre-planned maneuvers to achieve near simultaneous launch on the projected blues. The reds all turn on their radars to lock-on to blues just after their last turn towards the projected blue position. The reds launch radar guided missiles at their closest blue targets as soon as possible. Red 5/6 should pull delaying turns together then turn towards an intercept with US 1 (i.e., highest closure rate) once their radar acquires.

The blues split and turn towards the outside threats to take advantage of their longer range AMRAAM shots at each of outside red pairs. They support their launches until both outside reds are killed or until second shots are needed. In the baseline scenario shown, US2 achieves 2 kills with its first launches then turns towards reds 5/6 that have engaged US1 while taking its second shot at the surviving red 4. US1 will leave this second AMRAAM once it has acquired red 4, then pulls defensive maneuvers and countermeasures against the reds 5/6 missile launches while US2 completes red 5/6 kills.

The SUT gate multiplication factor was 5 and 15. The PE gate multiplication factors of 3 and 5, PE designs for Vogel and Hungarian based association, expected probability of false tracks, expected probability of detection and confidence ID updates.

Tier 1: Similar to Tier 0, the simulation for Tier 1 was run from time periods 1 to 329 with an interval of 1 time period.

Tier 2: The simulation for Tier 2 was run from time periods 1 through 329 at an interval of 1 time period.

5 Design of Experiments

When employing DOE test-planning methods, one issue that can arise is the complexity involved in designing efficient test plans if there are many independent variables (or "factors", the term used in the

DOE literature) whose effects on the DF process under test want to be known. Using traditional DOE experimental designs, the number of runs that have to be made will grow exponentially when the number of factors is large, and the number of "levels" (specific value settings of the factors) is large ; these go as the number of levels raised to the number of factors, or L^F. This exponential growth is associated with the type of experimental design being employed, called a "factorial" design, which not only allows the so-called "main effects" to be discerned from the experiments but also what are called "interaction" effects, where knowledge is gained about the effects on the metrics of interest due to interacting effects among the factors. If the desire to learn about the interaction effects is relaxed, using a type of experimental design called a "fractional factorial" design, the severity of the exponential growth is lessened but can still be an issue to deal with. Although we are still studying the strategies for and efficiencies of large factor-many level experimental designs, we are now employing a phased approach as shown in Figure 19, where we use the fractional designs initially as a screening step to determine those factors which are most influential on the metrics, and then the factorial designs to better understand the main and interaction effects of the key variables and, if necessary what are called "response surface" methods to understand the broad effects of the factors across the levels of interest for the application.

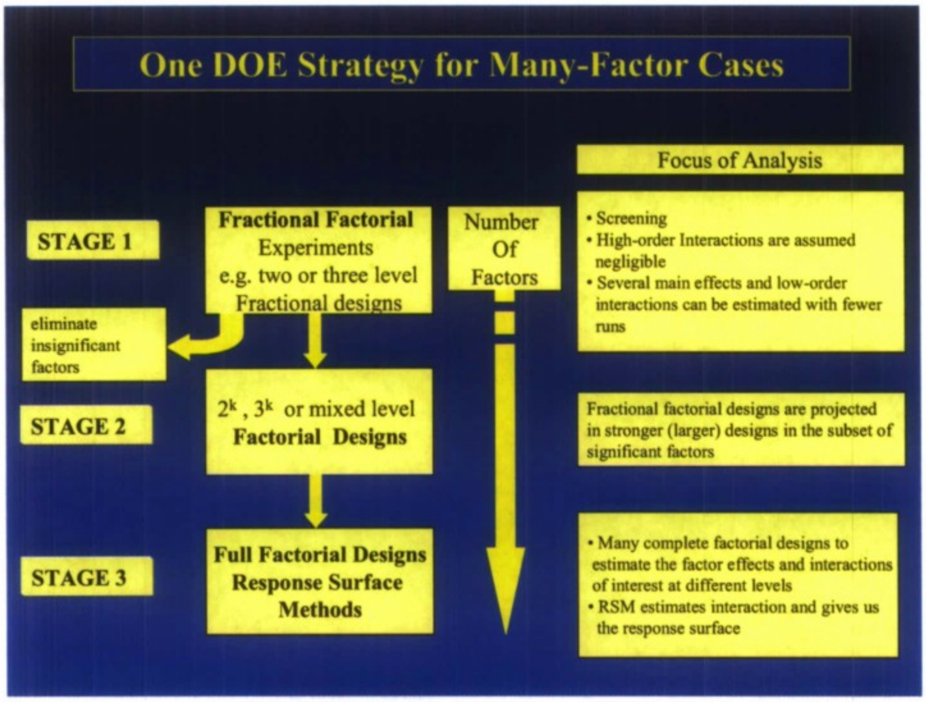


Figure 19. Phased experimental design strategy to deal with many factors and levels.

Perhaps the most important aspect of this formalized approach is that the post-test analysis procedures, generally falling under the title of "Analysis of Variance" or "ANOVA" procedures, allows the assessments of the results to be done with statistical significance. That is, assertions of the type that "the hypothesis that there is an effect of factor X on metric Y can be rejected with 95% statistical confidence" can be made as a result of the combined utilization of DOE test designs and ANOVA analysis methods. We planned a Design of Experiments (DOE) scheme for the PE MoPs. We conducted these tests on Tier 0, Tier 1 and Tier 2. We decided on the following factors to setup the DOE:

- **Scenario Factors (Fixed):**
  - Offensive Sweep 2vs6 Air-to-Air
- **PE Factors:**
  - Design (Association)

    Vogel Approximation (PE 1), and

    Hungarian based association (PE 2)

– Gating Factor: 3 and 5

- **System under test (SUT) Design Factors**:

  – Gating Factor: 5 and 15

So this yields a 2^k or 2^3 full factorial design. We used MINITAB to perform the DOE runs. The full factorial design details are as follows:

Factors: 3

Levels: 2

(A) SUT Design Gating Factor

(B) PE Gating Factor

(C) PE Design

Base Design: 3, 8

Runs: 80

Replicates: 10

Blocks: 1

Center pts (total): 0

All terms are free from aliasing. The factors and interactions that are significant for various MoPs are denoted by '**S**'. Table 3, Table 4 and Table 5 show the summary of the DOE run results for Tier 0, 1 and 2 respectively. In addition to these DOE runs, we ran another set of full factorial runs to see the effect of communication tiers on the various MoPs. We added another factor, (D) Tier, with two levels: Tier 1 and Tier 2. Table 6 shows the significant factors and their interactions for the various MoPs. The detail DOE charts are given in Appendix A, B, C and D.

Table 3. Tier 0 DOE run summary

		A	B	C	AB	AC	BC	ABC
Track 1 to Truth Radar	Aspect Consistency							
	ESM Consistency							

	TTLE Consistency							
	Threat Consistency			S				
Track 2 to Truth Radar	Aspect Consistency	S						
	ESM Consistency	S						S
	TTLE Consistency	S			S	S		
	Threat Consistency	S						
Track 1 to Truth ESM	Aspect Consistency	S	S					
	ESM Consistency							
	TTLE Consistency	S			S			
	Threat Consistency	S					S	S
Track 2 to Truth ESM	Aspect Consistency							
	ESM Consistency	S	S					
	TTLE Consistency	S			S			
	Threat Consistency	S						
Track 1 to Truth IRST	Aspect Consistency							
	ESM Consistency					S		
	TTLE Consistency							S
	Threat Consistency							
Track 2 to Truth IRST	Aspect Consistency							
	ESM Consistency							
	TTLE Consistency							
	Threat Consistency				S			

Table 4 . Tier 1 DOE run summary

		A	B	C	AB	AC	BC	ABC

Track to track	Aspect							
	ESM							
	TTLE							
	Threat		S					
Track 1 to truth	Aspect			S				S
	ESM			S				
	TTLE							
	Threat							
Track 2 to truth	Aspect			S				S
	ESM						S	
	TTLE	S				S		
	Threat	S						

Table 5 . Tier 2 DOE run summary

		A	B	C	AB	AC	BC	ABC
Track to track	Aspect				S			
	ESM							
	TTLE						S	
	Threat						S	
Track1 to truth	Aspect		S			S		
	ESM		S	S	S	S	S	
	TTLE		S	S			S	
	Threat					S		
Track 2 to truth	Aspect			S				
	ESM			S				

	TTLE							
	Threat			S			S	

Table 6 . Inter Tier 1 and 2 DOE run summary.

		A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD
Track to track	Aspect															
	ESM		S	S	S					S	S					
	TTLE				S											
	Threat		S											S		
Track1 to truth	Aspect	S	S	S	S	S	S			S	S		S	S		
	ESM		S		S					S	S					
	TTLE				S					S			S			
	Threat				S										S	
Track2 to truth	Aspect			S					S							
	ESM			S	S						S			S		
	TTLE				S											
	Threat				S											

The concept of analyzing the design for L4 design is really difficult. According to the Countermeasure logic implemented in Section 2.3, the estimated tracks generated will be significantly different than the truth track. The countermeasure logic suggests that jam the radar signal from the hostile platform, until the hostile reaches the cross-over range (S/J =1). After that the ownship changes its current trajectory to get out of the FOV of the hostile platform. Due to this countermeasure implementation the true threat will be different that the estimated threat and will hence forth have different tracks. To check for the change in S/J consistency over the truth and estimated tracks and the DOE charts are shown in Appendix E.

The statistical design shown here deals with very small number of factors. But in case of large number of factors we will need humongous number of runs to analyze the statistical design under study. We need to change our approach or find a new method to handle the increasing number of factors. In following section we have elaborated large factor DOE (Design of Experiment).

6 Large Factor DOE

Here we are trying to achieve a tradeoff between the number of factors, their levels and the cost of carrying out the experiment. As the former increase, the cost increases. In a full factorial design we analyze all the factors at all levels which have large number of runs. For example in a 2 level full factorial design with 5 factors a minimum of 32 runs need to be carried out. This makes it impractical to carry on with such an experiment. Generally this is the major reason experimenters opt for a fractional factorial design where information is obtained very easily and in fewer runs. But it is a highly confounded design. A lot of information is lost because of this confounding. As interactions are confounded with the main effect it becomes very unclear as to which is the factor that is responsible for the effect.

Goals of large factor design In order to overcome the shortcomings of a fractional factorial design, certain designs tend to the large number of factors and are economical and produce optimal results in very less time. The goals of such designs can be listed as:

- Lower number of trails with variations of multiple factors.
- Finding the separation of effects due to individual factors and interactions.
- Keeping a vertical balance

These designs have fewer runs and some of them even incorporate variations in them. They try to keep main effects separate from the interactions and find the influences of each of these on the experiment by solving them individually. They also try to maintain same number of runs at all levels of each factors. From some of Taguchi methods to other individual ones achieve these goals with ease where they solve right from 14 to 100 factors. The following subsections discuss these methods in detail.

6.1 Latin Square Design

A Latin Square design is used when there is a need to compare the treatments and to control up to 2 known sources of variation. As the other designs these were also first used to design agricultural experiments (Figure 20). The fertility trends were seen to run up and down and across the field. In such a scenario if there were 4 fertilizers used, then the field would be divided into 16 smaller plots by running 4 horizontally and 4 vertically. So when a Latin square design is used in such a scenario, the Latin square design will allocate the four different types of fertilizers in these plots with each type occurring only once in a row and column.

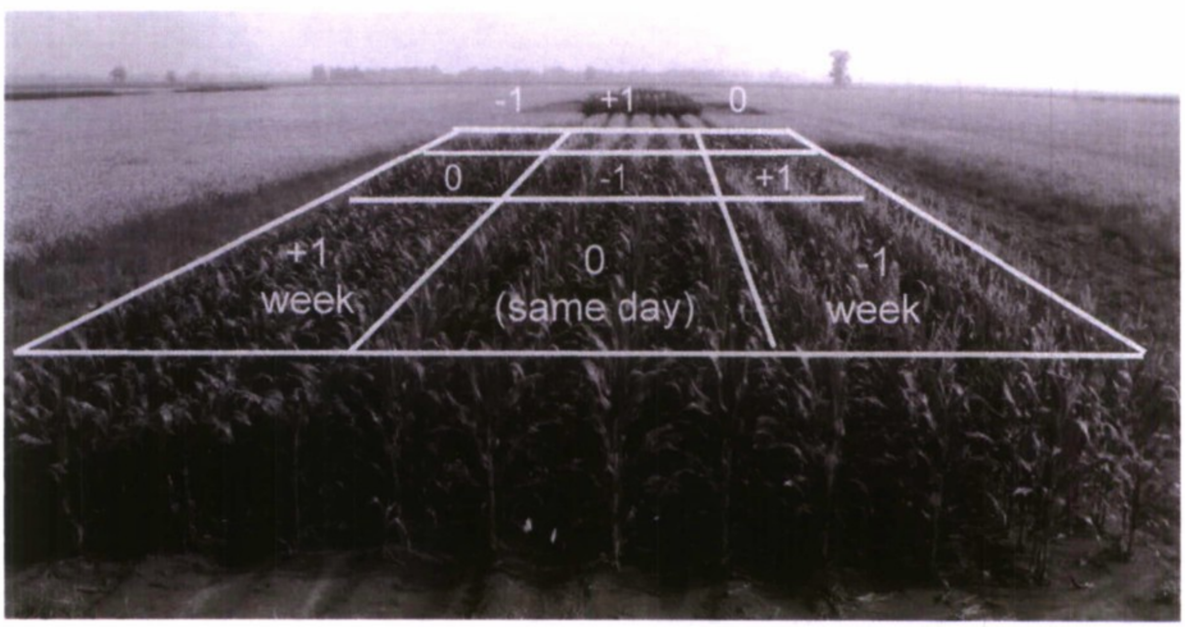


Figure 20. Example of a Latin square design applied in the field of agriculture.

In the early 70's researchers such as Finney [28][29], Federer [30], Freeman [31][32][33] and Addleman [34] explored on the concepts of generalizing the Latin Square designs and called them F square designs. Hedayat and Seiden [5] studied this concept in detail. In F square designs instead of the number appearing once in a row and a column, the number appears same number of times in the rows and columns. So these designs are explored to develop an understanding of them.

In a Latin square design the domain is needed to be known before hand. The factors of interest have more than two levels. These factors are also known well before hand. Each of these factors appears only once in each row and column. One of the most striking feature of such a design is that there are negligible or no interactions at all. It is one of the more complex designs. But it provides good results by making sure that the main effect of one factor does not bias the main effect of other factors. Unconfounded main effects are also derived in this design. Nuisance factors are used as blocking factors. The blocking is carried out in order to randomize the design. Lindquist argues that in a single Latin square design, the main effect is confounded with the interactions of the other two factors and also with the triple interaction. He also stresses that the residual of such a design is of ambiguous nature. The error is much lesser than that in a randomized block design (RBD) because there is a blocking factor used. The most common sizes of this design are 5x5 and 8x8, where a 5x5 design is shown in Table 7.

Table 7. A 5X5 Latin square design.

A	B	C	D	E
B	C	D	E	A
C	D	E	A	B
D	E	A	B	C
E	A	B	C	D

These designs have also been used in the field of medicine for cross over trials (suggested by Armitage and Berry [35]). This concept was further studied by Clark O Neil *et al.* [36] where they investigated 10 products on each of the patients. Since it needed to be orthogonal, 10 patient sets were chosen and a complete analysis was carried out. The benefits of using this design in a cross over trial were discussed in the paper by Kramer and Glass [6]. This design has also been used in the field of psychology, though very rare. In a paper by Gatio [7], the author discusses the inherent defects in the significance tests of a Latin square design when interactions occur under four conditions namely:

- One random variate model
- Two random variates model
- Three random variates model
- Four random variates model

Significance tests were carried out on all these conditions for all the effects and their interactions. The author concludes that before using this design, the user must get familiar with all the scenarios under which significant results are produced by this design. Also he says that in a psychological study, mainly the first two scenarios are prevalent. Information technology is also not very far from the other fields in the use of this design.

In the field of computer science, this design has been particularly used in compiler testing. Robert Mandl [8] uses this design to verify that evaluation of the operators on the enumerating values in ADA is correct even if these values were in ASCII code. The author illustrates this by using an example from a test. He concludes that the design provides comparatively great results in form of information for the amount of effort one puts in. He even feels it is quite a cost effective solution. The other common applications of these designs have been in the field of animal nutrition, insecticide field trials and even greenhouse effects. There are two variations of this design. But a detailed analysis is out of the scope of the thesis.

- ***Graeco Latin Square Design*** – It is a kxk design just like the Latin square design but the information gathered from this design is more than just a normal Latin square design. For example if a 3x3 design is considered one can evaluate the main effects of four 3 level factors in just 9 runs.

- ***Hyper Latin Square Design*** – It is also a kxk design similar to the Graeco Design but differs from it as it uses more number of blocking factors. For example if a 4x4 design is considered one can evaluate main effects of five 4 level factors in just 16 runs.

6.1.1 Advantages

- Several nuisance factors are handled with these designs either when they need to be treated separately or they should not be combined in to one factor.
- Numbers of runs are very small.
- At least two sources of variation are controlled.

6.1.2 Disadvantages

- It should be a square design as in the number of levels of each blocking variable must equal the number of levels of the treatment factor.
- A major assumption made by this design is that there are no interactions between the blocking variables and between the main variable and a blocking variable.
- The degree of freedom associated with the error term is relatively small for a small design.
- If the number factors are more, the design tends to get bigger and the error term associated gets lager.
- If there are any missing values then the design becomes statistically too complex.
- The following interactions cannot be evaluated:
  - Rows and columns
  - Rows and treatment factors
  - Columns and treatment factors

6.2 Plakett-Burman Design

During the initial stage of experimentation when there is minimal knowledge of the problem in hand, screening experiments are conducted in order to find the major factors in few runs. Until 1946, the most common design used to conduct these screening experiments was the fractional factorial design. But R.L. Plackett and J.P. Burman in their famous paper "The Design of Optimal Multifactorial Experiments" described a new economical and efficient design for screening experiments. The design was named after them. In an un-replicated fractional factorial design the number of runs is restricted as a power of 2 whereas in a Plackett-Burman design the number of runs is treated as a multiple of 4, hence making it economical in obtaining the factors in fewer runs. This design is used only when main effects are of importance because the main effects are highly confounded with 2 factor interactions. There is no defining relation because interactions are not identically equal to the main effects. They are resolution III design known as saturated main effect. For example just 12 runs would be needed for up to 11 factors.

Table 8 . The six-factor Plackett-Burman Design used in the first conjoint study.

Factor No/ contrast	A	B	C	D	E	F					
	l_1	l_2	l_3	l_4	l_5	l_6	l_7	l_8	l_9	l_{10}	l_{11}
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	-1	-1	-1	-1	-1	1	1	1	1	1	1
3	-1	-1	1	1	1	-1	-1	-1	1	1	1
4	-1	1	-1	1	1	-1	1	1	-1	-1	1
5	-1	1	1	-1	1	1	-1	1	-1	1	-1
6	-1	1	1	1	-1	1	1	-1	1	-1	-1
7	1	-1	1	1	-1	-1	1	1	-1	1	-1
8	1	-1	1	-1	1	1	1	-1	-1	-1	1

9	1	-1	-1	1	1	1	-1	1	1	-1	-1
10	1	1	1	-1	-1	-1	-1	1	1	-1	1
11	1	1	-1	1	-1	1	-1	-1	-1	1	1
12	1	1	-1	-1	1	-1	1	-1	-1	1	-1
l_i represents the i^{th} contrast and $l_T l_{11}$ consist of interaction effects											

There are 2 kinds of Plakett-Burman designs – geometric and non geometric. It is a geometric PB design when the number of runs can be depicted as a power of 2 if not it is non geometric. The design is distribution of equal number of pluses and minus in a column. For each of the n-1 columns, the design allows the contrasting of data by taking the difference between the averaged data opposite to these signs. In the geometric designs, the columns are orthogonal. But in a non geometric design, the contrast columns are mutually orthogonal but at the same time they can be correlated to contrast columns of the interactions. Due to this the major factors may not even show up on the radar making the analysis inaccurate. Even though there is such a complex alias structure, the design under some circumstances works and also estimates the interactions simultaneously which other designs are unable to do.

In order to analyze the complexity of the alias structure formed in this design, many authors have proposed various methods. During screening experiment, based on the assumption that only some factors are the cause of variation in the experiment (Box and Meyer [37]) and that for an interaction to be significant, the corresponding main effects also should be significant, Hamada and Wu [38] proposed one such method in which they contrast the main effects and two factor interactions orthogonally to the ones found by standard methods. But their design is limited to 2 factor interactions. Some studies by Hynen [39] showed that due to aliasing in the design unwanted two factor interactions would appear if the major factors were more in number. Box and Meyer [40] also proposed a method which employed Bayesian methodology in order to determine if the factor is a major factor or not. In the paper by Tyssedal and

Samset [9], the authors try to makc the PB design more robust design by trying to overcome the shortcomings. They provide an alternative method for analyzing the design by understanding the alias structure and by finding eonsisteney between the analysis and the projeetion properties of the PB design [10].

The Plackett-Burman designs have seen light in various fields of researeh and practical applieations for sereening. In the paper by Devos *et al.* [11] thc authors use this design to ealibrate partial least square regression model which was being used to test six polyclinic aromatie hydroearbons and compare it with the ones calibrated using eollinearity. They eonclude that the results obtained by using this design arc optimal. This design can casily solvc more than 27 faetors in just 28 runs. Plackett Burman design has also been used in the field of Bioteehnology to find the effect of the medium in xylanase produetion using a 12 trial design [41] (Li *et al.*).

Tyssedal and Samsct [9] use a 12 run 9 response non gcometric PB dcsign to makc an injection moulding cnvironment, where 15 to 20 variables are eonsidercd while produeing a new plastie component. It is seen to be a more cost effeetive solution and also eyele times are reduced. The design is also used to obtain signifieant parameters rapidly and objeetively in a thcrmal proeess which synthesizes eompounds.

Anothcr application of these designs has been in agrieulture to find the main extraction faetor that affeeted the yield and quality of peetins in ehieory root. The authors deeided to use a two level design with 17 factors and 20 experiments because they faced a problem as there were too many faetors and werc unsure of the settings which produced optimal results [13].

One of the main applications of thcse designs is in the simulation experiments. They help in screening out the important faetors used in simulating the design. They are used in setting up and analyzing the eomputer arehiteeture simulation cxperiments. The solution was proposed to use sueh a design as the authors observed that the proeessor simulation does not follow any partieular statistical method and thc results are not in any eonfidenee levels. Taking this eoneept further, Vanderster *et al.* [15] use this design to optimally select the parameters for a knapsack metascheduler which provides a way to systematically allocate policies on a computational grid.

A Plackett Burman design generally produces a saturated design. In order to solve these saturated designs a new method called Fixing Effects and Adding Rows (FEAR) has been discussed in a paper by Heyden *et.al.* [14] Here the authors have described a model by adding zero effects rows to the model matrix after which the largest estimated effect is fixed in order to examine the factor effects accurately. This method helps in estimating the effects of the factors. With a set of data values, a comparison between FEAR and the conventional Multiple Linear Regression method was carried out and from the results it can be seen that the new proposed method performs better than the conventional method of regression because the main effects that are significant are estimated more accurately in FEAR than when compared to Multiple Linear Regression. If the PB design is complete then the error is very less even if not they are insignificant and the design produces satisfactory results [14].

6.2.1 Advantages

- Economical for detecting large main effects (assuming that all the interactions are negligible in comparison with few important main effects).
- Due to the confounding the negligible impact is averaged and information may be obtained about significant interactions.

6.2.2 Disadvantages

- Due to confounding, the presence of a large interaction may distort the effect of an individual factor.

6.3 Split Plot Designs

Split Plot designs are one of the most robust designs found. Though Taguchi's approach is also robust; the size of the experiment is large and needs a large number of trails. Recognizing this drawback, Box and Jones [42] suggested the use of Split Plot designs, which save the number of runs and even provide additional information Split plot designs are an extension for randomized block designs. These designs are used in an industrial application when there are moments where various factors (processing and set up) and levels are difficult to understand or the experiment is more expensive or it is a very laborious and time consuming experiment. To overcome these problems, the re ordering of runs is carried out which results in a Split Plot Design. Instead of carrying out experiments and then reordering it into a Split Plot, the design can be used in the first place. Many a times this goes unnoticed. In a paper by Kowalski and Potcner [16], there are guidelines to recognize a Split Plot diagram. They say that these designs have three main characteristics:

1. The levels of the factors are not random and are reset after each run.
2. For each factor the size of the experiment varies.
3. The random assigning of the treatment combinations to the experimental unit is not allowed.

Though large in size these are cost effective design as shown by Webb *et al.* [44]. Also in the paper by Bisgaard [43], there is a cost model for such a design along with examples. These illustrations also show a reduction in the runs and additions in the information obtained.

There are 2 levels or as Goos and Jones [45] say strata in these designs. The experimental runs are divided amongst these strata. The upper stratum is made up of whole designs and the lower stratum consists of subplot runs. The whole plots are a group of runs where the factor which is difficult to change remains constant. This nomenclature of the strata came to existence as these designs were like many other designs used in agriculture originally. For example; the use of the fertilizer or an irrigation method as one of the factors that can be applied to the large sections of land called whole plot came to be known as the whole plot factor and the factor associated with variety of seed to be ploughed in various sections of this land by splitting it into sub plots came to be known as the subplot factor. Such a design is used only when there are many stages in the experimental design. There can be reordering of designs within these strata which give rise to split- split- plot diagrams. The most common example used while understanding this method is the production of cheese, a case study by Schoen [46]. There is not much literature to understand the details of a three stratum experiments. However, details on how to design and analyze two level factorial and fractional factorials are provided by Bisgaard [43].

There are a few additional assumptions that an experiment needs to meet before the use of Split Plot Design. They are:

1. There are two or more independent variables such that one is a non repeated measure treatment or between-block treatment and there is at least one repeated, or within-block treatment.
2. The number of combinations of treatment levels is greater than the desired number of observations within each block.
3. If repeated measurements are used on factors, then each block will consist of only one factor; if not there will be more factors.
4. The sequence for administering the repeated measures levels in combination with each level of the non repeated measures treatment is randomized independently for each block.

The Split Plot Design is one where a factor is subjected to all levels of some treatments but only one level of the other treatments. These designs combine both the features between the plots as well as within the plots. This feature helps in saving the number of runs and thus a large amount of information is gathered very efficiently. The first level factors are randomly assigned to the whole units depending on the whole unit design. The second level factors are assigned to the sub units randomly within each whole unit according to the rules of Randomized complete block design (RCBD). Thus the entire design is randomized. It would not be too harsh to say a good understanding of the domain is needed in order to use this design.

Table 9. Example of a 3x2 Split Plot Design.

A1	A2	A3
B1	B2	B1
B2	B1	B2

Table 10. Randomization of the 3x2 design shown above.

A1	A2	A3
B1	B1	B2
B2	B2	B1

When an ANOVA is conducted on such a design, it is seen to have two error terms. The first error corresponds to the pooled variation between the factors within the groups and the other corresponds to the pooled interaction of the treatment with factors in each block. This happens because there are two separate randomizations that occurred when the experiment was run [16]. According to Kowalski and Potcner [16] in a Split Plot Design one needs to be sure that there is a true replication in the whole plot factor.

From the varied literature, the applications of this design can be seen in

1. Experiments in which the each factor has a need for different number of experimental units.
2. Experiments where one factor needs to be more sensitive than the other.
3. Experiments where there is a need to introduce new factors unexpectedly.

6.3.1 Advantages

- The sub plot's treatment factor and interactions are tested to a generally high sensitivity than the whole plot because the variance in the former is much higher than in the variance in the latter.
- Experiments with a large number of whole plots and lesser number of sub plots can be conducted in a single experiment.
- Factors may be added with minimal additional cost.
- It is a design where whole units are subjected to repeated measuring and these repeated measures are the sub plots.

6.3.2 Disadvantages

The design is very robust however there are a few shortcomings which are listed below.

- The existence of two error terms causes complicated analysis as there are many standard error comparisons.
- The sensitivity of the whole plot is poor as there is high variance and few replications associated with it.

6.4 Central Composite Design

Generally during the Design of an Experiment, mean models are estimated by assuming that a design is homogeneous. Central Composite Designs are the second design type discussed so far which are used in screening experiments. It was proposed by Box and Wilson in 1951 [47]. These designs are an extension of the factorial designs. When either the full or fractional factorial designs are embedded with centre points and axial points or star points; a Central Composite Design is formed (Figure 21). The centre points are experimental runs in the design whose values correspond to the median values in the factorial design and the design is usually replicated along this centre. The axial or star points however are the points which aid in the rotation of the design by adding curvature to the design, mainly by including the upper and lower median values of the two factorials.

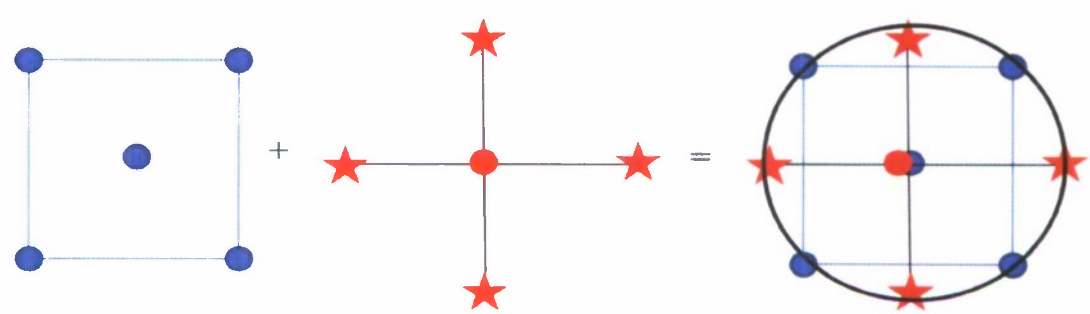


Figure 21. Formation of a central Composite design.

A second order central composite design is an alternative design to a 3 level fractional factorial design or a 3^k design. As k increases the design size is greatly reduced in a Central Composite Design (CCD). As a known fact such a design contains twice the number of star points as the number of factors present in it. Like in a factorial design one can choose the value for high and low levels, in a composite design the values are represented by these star points. There are different kinds of Central Composite Designs based on the position of the star points. They are:

- Circumscribed composite design
- Inscribed composite design
- Face centered composite design

A Circumscribed Composite Design requires 5 levels and the star points define a new high and low values. With these values as a limit, if a factorial design is created, then it is an inscribed composite design which also requires a 5 level factorial design. In a face centered composite design there are 3 levels of each factor. The star points are in the centre of the faces of the design. The inscribed and the circumscribed designs are rotatable. But a face centered design is fixed. The largest amount of space is covered by circumscribed designs and the least by inscribed design.

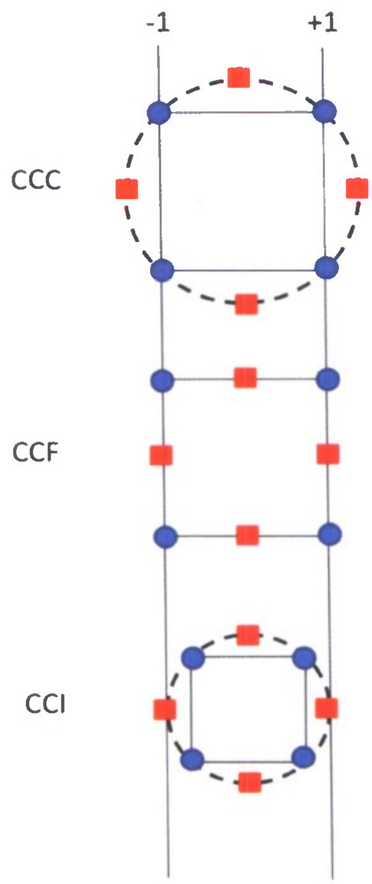


Figure 22. Different Kinds of Central Composite Design.

Many authors have used this method to solve up to 120 factors and tested in evaluating the weapon system. In the paper by Sanchez and Sanchez [17] the drawbacks of a large full factorial design is discussed. Owing to this discomfort in solve a highly fractionized factorial design is proposed that solves large designs. These highly fractionated designs are the central composite designs. The authors also show that their designs is double the size when the factors are in the range of 30-32 and 53-64 but they required few centre points. With the number of factors considered, there is a variation in the designs. Some work has been carried out in order to recognize these variations. In a paper by Li *et al.* [12], the authors compare the designs based on the variances for both rotatable and non rotatable designs. The numbers of factors considered are between 6 and 10 with the consideration to axial points. They conclude that the CCD based on a resolution 5 design performs really well and in fewer runs.

Bjorkman and Zeius [18], discuss the application of this design in their paper. The challenges faced by the military in a decade to come, with respect to the testing, have been well recognized by the authors in this paper. Further the authors propose various process capabilities based testing methods such as central composite designs in order to overcome these challenges. The authors justify the use of this method by saying that this design would solve large factors with low variance.

This design also finds application in the textile industry. In a paper by Kothari *et. al.* [19] the authors use the design in order to understand the factors that affect air jet texturing which forms neps which reduces the quality of the yarn. Similarly, the designs have been used in the manufacturing industry to find the optimal factor for the flux cored arc welding and to optimize the design. Similar to the former paper the authors recognize 4 factors which are used to optimize the design.

The Central Composite Designs is a response surface methodology and hence the entire surface is under the study. New work has been carried out by Hader and Park [48] on these designs to make their slopes rotatable which reduce the variances in these designs. This rotatability is achieved by adjusting the distances between the axial points. Sometimes in certain experiments, the factors cannot be changed easily. But Kowalski *et al.* [20] have worked towards developing a better understanding of this problem and providing a solution to it. Kowalski *et al.* [20] modified these designs to accommodate a split plot structure. Here the authors modified the Central Composite Design which helps in the estimation of different models based on their mean and variance under a split plot structure taking it a step further. These designs have also been used to determine the property and structure of certain epoxies. In the waste water treatment, these designs have been used to optimize the parameters. Such is the diversification of the use of these designs.

Non-central composite designs are an attractive alternative to the central composite designs when the design is asymmetric by the shift of interest after conducting an initial 2 level design. In a paper by Robert Mee [21], the author discusses this alternative and gives instances to support his study.

6.4.1 Advantages

- The designs can be run sequentially.
- They are very efficient as they provide a lot of information on the variables in fewer runs.
- The experimental error is also determined in very few runs.
- The CCD's are very flexible.

6.4.2 Disadvantages

- It's a resolution 5 design which is higher than some other designs.
- The surface plots are not rotatable.
- Sometimes the interactions between some variables and square terms are lost.

6.5 Taguchi Methods

Some of the Taguchi methods are also used to solve experimental designs. These are designs pioneered by Dr. Genichi Taguchi, which have helped in process improvement by improving the productivity. This method is a philosophy by itself. In itself it has 2 main doctrines:

- To decrease the inherent variation in any process
- To develop a strategy in order to carry out the above stated doctrine.

Some of these strategies could be to identify which of the parameter in the process will help in improving the strategies. It could also be done by identifying an alternative which will yield the same or better results. By far it is one of the most robust designs known in the industry. The noise factors and cost of failures are incorporated in the designs which ensure customer satisfaction. These designs are used for optimizing the design of performance, quality and costs of any equipment. Taguchi method for designing an experiment is used mainly in manufacturing processes. Figure 23 below gives a pictorial representation of the design.

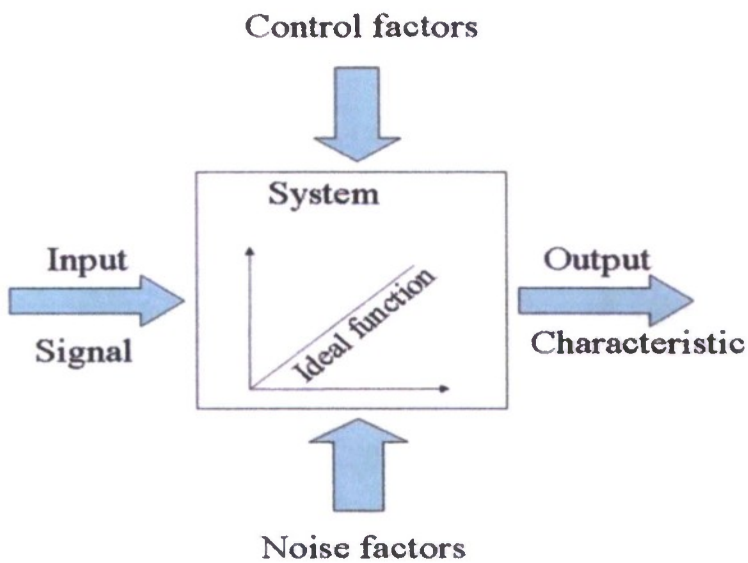


Figure 23. Taguchi System Representation.

It is one of the highly debated methodologies [22]. In [22] T.N. Goh discusses the issues as to why some authors support this method and some oppose. He says the argument lies in the technical merit of this method. Some authors say that this method has no alternative that could predict the improvement. But the believers say it is easy to be used even by a person who does not have in depth idea about the mathematics of the method. Extremist also argue that the variations in the environment are not considered and the design can "hide" the requirements for optimization of the response and minimization of the variance in the design making it counterproductive. But these designs have been widely used and the results are hard to ignore which makes it a valid design for consideration in the thesis. Mainly Taguchi designs can be broadly classified as Orthogonal Arrays and Linear Graphs.

In an orthogonal array experiment the columns of the independent variable are "orthogonal" to each other. They are often used when there are a number of control factors in the experiment. These designs are fractionated factorial designs. Orthogonal arrays have to be defined in terms of the number of factors considered, the levels of the factors and the specific interactions of interest.

Table 11. Orthogonal array eleven two level factors.

Experiment	Column

Number	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

When there is a need to assign the factors in an orthogonal array, linear graphs are used. These are substitutes for triangular table, also a Taguchi method. As mentioned in [27] by the author, the graphs have nodes and lines. The numbers present beside the vertices and the edges correspond to the columns in the orthogonal array. The vertex of the graph shows the factor and the edge shows the column of the interaction between the connected vertices.

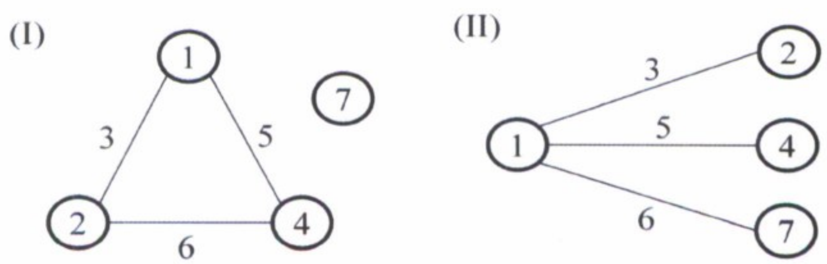


Figure 24. L8 linear graph.

The operational steps of a Taguchi design is depicted in the flowchart below (Figure 25)

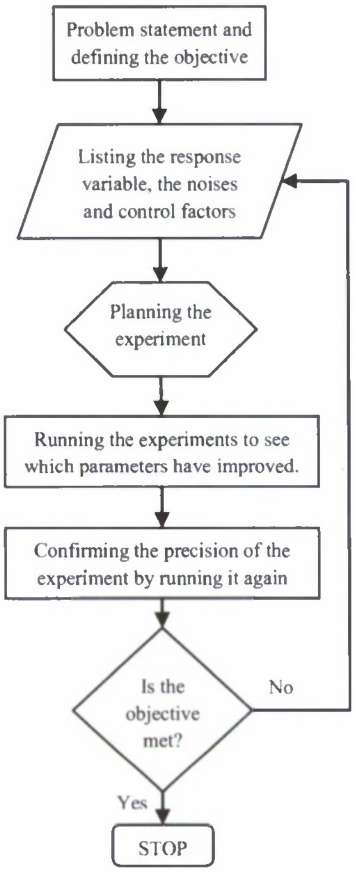


Figure 25. Operational steps of the design.

In the paper by Antony *et. al.* [23], the authors have used the Taguchi methods in an automotive industry to develop a new coil. An experiment is designed with 16 trials to study 14 parameters with one interaction. The authors [23] follow the steps of the Taguchi method diligently and come up with optimal settings for the design parameters that are very important in making this coil. In conclusion they talk about the effect of such designs in solving large problems easily in industry.

6.5.1 Advantages

- Does not consider specification limits but the quality of the system.
- Error is resubmitted.
- Uses the noise factors in the experiment.

6.5.2 Disadvantages

- Orthogonal arrays are used without a thought.
- Interactions are not considered.
- There is no modeling but analysis results in the answer.

6.6 Analysis of the Designs

6.6.1 ANCOVA

In the former section of this chapter, the different methods that are used to solve large factor problems are learnt. Generally, after conducting the experiment, an ANOVA is conducted to find out the variances in the design. ANCOVA is a method that takes the analysis of variance a step further. ANCOVA can be pictorially defined as:

ANOVA

+

Linear Regression

=

ANCOVA

Figure 26. ANCOVA pictiorial depection

From Figure 26 one can know that ANCOVA has two components embedded in it. Along with a regular ANOVA, ANCOVA includes linear regression. In order to understand this method better, covariance needs to be well understood. In simple terms covariance is the degree through which two variables vary together. A covariate is the source of independent variation that affects the response variable but was unknown in the beginning of the experiment. This is helpful because, it helps in reducing the unknown variance in the design which aids in estimating the means of groups more precisely.

Knowing this, ANCOVA can be defined as the method that compares mean values of response variables between groups when response variable co varies with other continuous variables [24].

6.6.1.1 Formulation

Mathematieally ANCOVA might be represented as

$$y_{ij} = \mu + \alpha_1 + \beta(x_{ij} - x) + \varepsilon_{ij}$$

Where

y_{ij} = jth replieate of ith level response variable

μ = mean value of response variable

$\alpha_1 = \mu_1 - \mu$ = differenee between the means

β = combined regression eoeffieient

x_{ij} = eovariate of the jth replieate observation from the ith level of a faetor

x = mean value of covariate

ε_{ij} = unexplained error assoeiated with jth replieate observation from the ith level factor

6.6.1.2 Application of ANCOVA

Since there is a covariate used in this method, the residual variation is removed. The method hence can test whether certain factors have effect very easily. It is known to statistically more powerful than a one way or even a two way ANOVA, since it accounts to some variability in the designs. Adding a covariate to ANOVA reduces the degrees of freedom of the design. But it is a dependent on there being a correlation between the covariate and the response variable. Adding a covariate which accounts for very little variance in the dependent variable might actually reduce the statistical powers if not the power of the design is very high.

6.6.1.3 ANOVA for ANCOVA

An ANOVA table of ANCOVA is represented as shown in Table 12.

Table 12. ANOVA table for ANCOVA

Source	Df	MS	F-Ratio
Factor A (Adjusted)	(p-1)	$\frac{SS_{A(adj)}}{(p-1)}$	$\frac{MS_{A(adj)}}{MS_{Residual(adj)}}$
Residual (Adjusted)	p(n-1)-1	$\frac{SS_{Residual(adj)}}{p(n-1)-1}$	
Total (Adjusted)	pn-2		

By using this analysis a statistical control of the error is obtained which is a strong point of the design. There is no upper limit in the number of factors that can be considered by this design. But a drawback of the design is that assumptions need to be met and there needs to be correlation between the covariate and the response variable.

6.6.1.4 Assumptions

The assumptions are as follows:

- Normality
- Homogeneity of variances
- Independence
- Linearity
- Covariate values should not be different amongst the group
- Fixed covariate
- Homogenous slopes - These have to be tested for compulsorily

Thus the designed can be summarized as an extension of the ANOVA where covariate is included which helps in increasing the statistical power of the experiment with the only limitation being – assumptions need to be met.

7 Classification and Recommendation of Designs

Foremost idea of the thesis has been to develop an understanding of the different methodologies that are used to address the issues of large number of factors. The mathematics behind the design, their complexity and number of factors they can solve, their advantages and disadvantages and the application of these designs has been discussed. But this has been mainly through literature survey. Taking a step further, these designs have been analyzed and classified based on the general understanding. Based on these classifications a recommendation for the use of these designs has been provided. This recommendation can be used as an aid in the selection process, thus serving as a guideline.

7.1 Classification based on Advantages and Disadvantages

After a thorough literature survey of various cases, the advantages and disadvantages of each design is listed in the following manner (Table 13):

Table 13. Advantages and dis-advantages of all designs.

Design	Advantages	Disadvantages
Latin square design	• Several nuisance factors are handled with these designs (either when they need to be treated separately or when they should not be combined in to a single factor) • Fewer numbers of runs. • The variations can be controlled in 2 directions. • The efficiency can be increased when compared to RCBD. • Economy of samples and ready analysis.	• Number of treatments should be equal to the number of replicates. • The experimental error is likely to increase with the size of the square. • Smaller squares have fewer degrees of freedom for experimental error. • The following interactions cannot be evaluated: ○ Rows and columns ○ Rows and treatments ○ Columns and treatments.

Plackett-Burman Design	• They give same info as $2^{(k-p)}$ resolution III design but with your fewer trials. • It is a sequential experimentation which is valuable. • By this design the variables can be enhanced. • It is a feasible design. • It provides robustness in the product. • It helps in intelligent decision making. • Helps in finding which variable can be used to change the system. • It does not have the power of 2 restrictions, since it is a 4N design making it more flexible.	• The amount of a priori knowledge of the experiment is important while using this design. • Aliasing pattern for such a design is very complex. • It is not a great idea to run large experiment or to depend on strategies that do not have the possibility of resolving complex relationships among factors with only few additional runs.
Taguchi methods	• Compress the amount of data required to carry out the experiments. • Give benefits of multiple simultaneous AB split tests. • Allows the testing of a few pages of elements all at once. • Requires far lesser data than a normal design. • Gives a robust design. • To achieve the objective the number of trails required to be carried out is very low.	• Though the size of the experiment is larger than a normal design, it is small when compared to some other designs. • Requires in depth knowledge of DOE. • In order to understand the results, high statistical knowledge is required. • The variable interactions are not considered in this design.

ANCOVA	• Power of the experiment is higher than the others. This is because there is a reduction in error variance. • By this method residual variation is reduced. • It is highly powerful when compared to 1 or 2 way ANOVA. This is because it has a greater ability to detect and estimate the interactions (within the group as well as between the groups) • There is availability of extensions to deal with measurement errors in covariates.	• It does not yield results if the assumptions are not met at least approximately. • The dependent and the independent variables should be linear in parameters. This further leads to the fact that there should be correlation between the covariates and the response variable. • There is an additional cost of introducing blocking factors. • The blocking factors that are highly correlated with the dependent variables become hard to find. • If the blocking factor is poorly correlated, there is a loss of power. • This design reduces experimental error by statistical methods rather than experimental methods.
Central composite Design	• The design gives highly accurate and strong results because the detection limits are lowered. • The design provides equal precision for fitted response at the points. • Such designs identify multivariable interactions. • The numbers of trails conducted to reach the	• They employ 5 levels for each factor, which is higher than some of the other designs. • At times such designs are not able to determine the interactions between the variables and the square terms. • The surface plots of such designs are not rotatable making them give different answers at different points.

	required conclusions are minimized. • Such a design also determines the different factor levels that provide optimum responses. • Also helps to determine the portion of the response that is insensitive to changes in predictor variables. • They fit non linear models. • These designs can be used to analyze data of any kind.	
Split Plot Designs	• The sub unit variance in such design is far lesser than the whole unit variance. Thus the sub unit treatment factor and interaction are generally tested with much higher sensitivity. • Such designs can carry out both the whole unit and sub unit analysis in the same experiment. • They follow a univarite design and have repeated measures in time (sub unit) carry out the whole unit analysis.	• Less precise than a fully randomized experiment. • Many designs have too few degrees of freedom to give good estimates of the main plot variation. • Analysis becomes more complex in cases such as missing data value, existence of covariates or while carrying out regression analyses. • There are 2 kinds of errors and hence there are many standard error comparisons. • High variance and few replication of the whole unit lead to poor sensitivity on whole unit factors. • Presentations of results are harder.

From the list above certain inferences of the designs are derived. It can be seen that the designs have certain common features. The designs have been grouped further based on these features into classes.

When just the advantages are considered the designs can be broadly classified as:

- Class 1A- Economical
- Class 2A- Large amount of information gathered
- Class 3A- Number of runs/trails are low
- Class 4A- Control of variations
- Class 5A- Aid in decision making
- Class 6A- Solve Non Linear Models

Though these are self explanatory, a brief description of them is discussed. As discussed earlier, these designs are generally economical. But this is a very subjective interpretation. Depending on the objective of the experiment and the situation of use, some of the designs might be more feasible than others. Mostly, a design is economical when its usage at least achieves a breakeven. Some of the designs take an additional stride and solve even non linear models. Most of the designs discussed are robust in a certain way, some based on their ability to control variation in the process and some based on the amount of information gathered which is why they become such an important criteria for classification. If the later is achieved in the least amount of runs, the design achieves its efficiency. These designs can be used in decision making. How useful they are as an aid, becomes another selection criterion. Based on the above criteria, the following classifications of the designs are made (Table 14).

Table 14. Classification based on advantages.

Class 1A	Latin Square Design, Taguchi Design, Central Composite Design, Split Plots, Plackett-Burman Design
Class 2A	Plackett-Burman Design, Taguchi Design, Split Plots
Class 3A	Latin Square Design, Plackett-Burman Design, Central Composite Design, Split

	Plots
Class 4A	Latin Square Design, ANCOVA, Split Plots
Class 5A	Plackett-Burman Design, Taguchi Design, Central Composite Design.
Class 6A	Central Composite Design

In order to get a quick idea as to which of these advantages is the most lucrative to be considered while designing an experiment a bar graph is quantified.

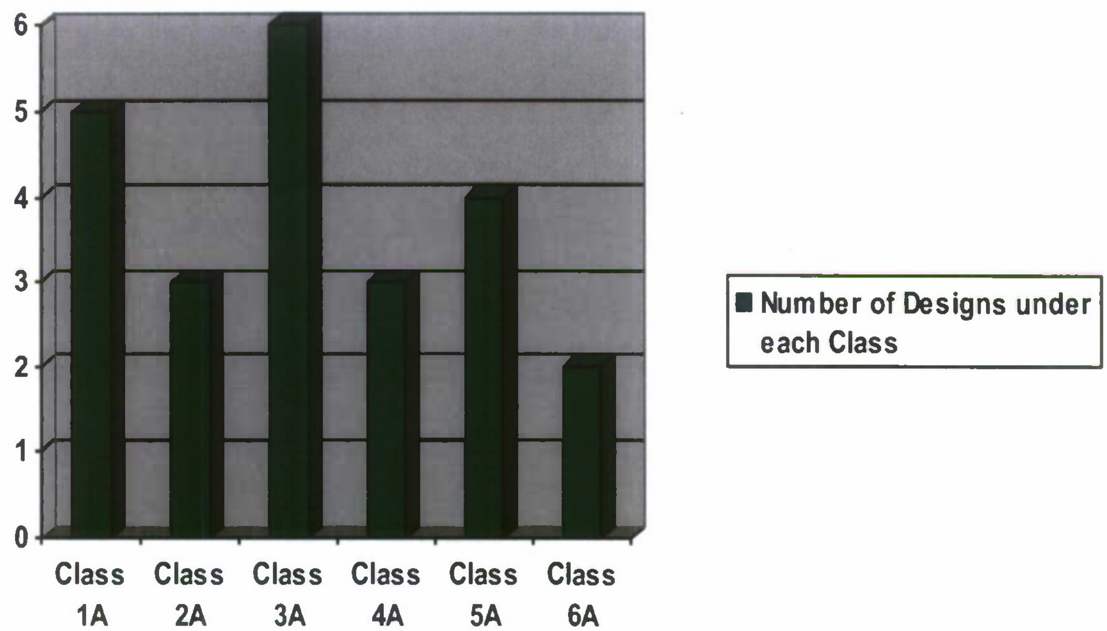


Figure 27. Number of designs for under each class/advantages

To give a fair idea of the designs, it is unsettling if only the positive are considered. Therefore, the disadvantages of these designs also need to be known. Based on the list of disadvantages, the designs were broadly classified as

- Class 1D - Assumptions for the design need to be satisfied
- Class 2D - Complexity of the design
- Class 3D - Prior knowledge of design needed is high/Domain should be known
- Class 4D - Limited analysis of the interactions

- Class 5D - High cost
- Class 6D - High Error
- Class 7D - Low degrees of freedom

As we already discussed while classifying advantages this is a very individualistic interpretation. As these designs are large, they can get very complex as the factors increase making it hard to understand. This intern leads to higher cost of the design and some of them are associated with a higher error term caused due to lower degrees of freedom. Most of the designs require a good amount of domain knowledge; the assumptions needed to be satisfied are quite a few. Designs like Latin squares limit themselves to the main effects and interaction analysis is not carried out. Hence these act as the parameters that measure the flip side of the designs. After classifying the disadvantages into classes, the designs are fit into these classes in the following manner (Table 15).

Table 15. Classification based on disadvantages.

Class 1D	ANCOVA
Class 2D	Plackett-Burman Design, Split Plots
Class 3D	Plackett-Burman Design, Taguchi Design, ANCOVA, Split Plots
Class 4D	Taguchi Design, Central Composite Design
Class 5D	Plackett-Burman Design, ANCOVA
Class 6D	Latin Square Design, ANCOVA, Split Plots
Class 7D	Latin Square Design, Split Plots

A bar graph drawn (Figure 28) will aid in quantification of these disadvantages while designing an experiment.

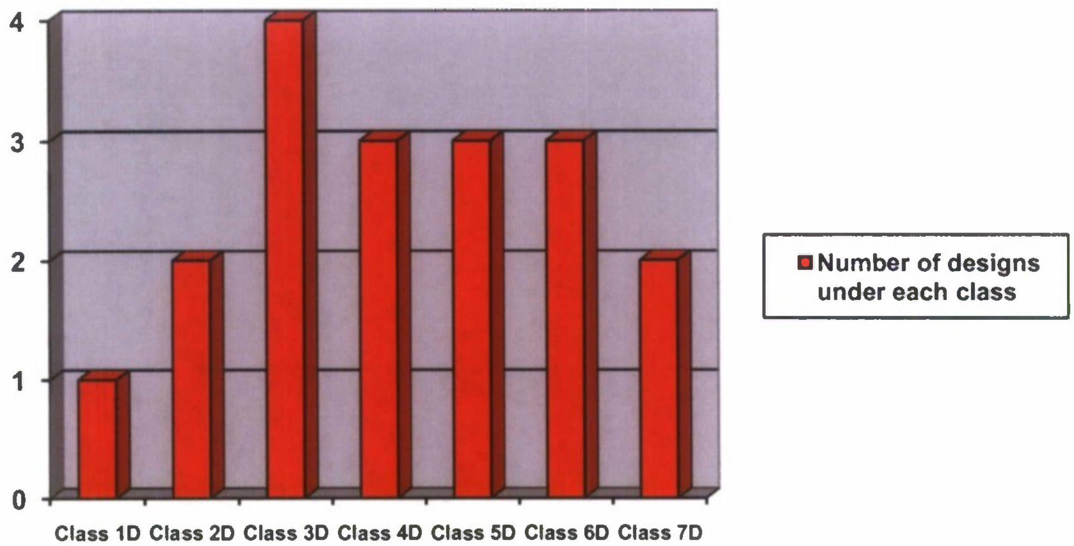


Figure 28. Number of designs under each class/disadvantages.

7.2 Classification based on number of factors and complexity of the design

As already stated, the number of factors a design can solve is one of the major reasons the research has been conducted. In order to achieve an optimistic result for this study, the classification based on the number of factors is carried out. This is a nascent step towards providing some kind of guidelines in order to achieve that tradeoff between the cost and the information used and got from these designs. The larger the design the cost is higher. So considering the minimum number of factors that can be solved with the least amount of effort the designs were classified as

- High ($x>15$)
- Medium($10<x<15$)
- Low($5<x<10$)

where x denotes the number of factors.

Having classified in this manner, the designs can be grouped into these classes and a graph is developed.

- High - Latin Square Design, Split Plot Design, Taguchi designs, CCD

- Medium - Plackett-Burman
- Low – Fractional Factorial Design

The Figure 29 below gives a quick update on the designs and aids in picking the most suited design for carrying out the experiment. The numbers of factors vary as high medium and low and the complexity involved in solving the design varies from low to high.

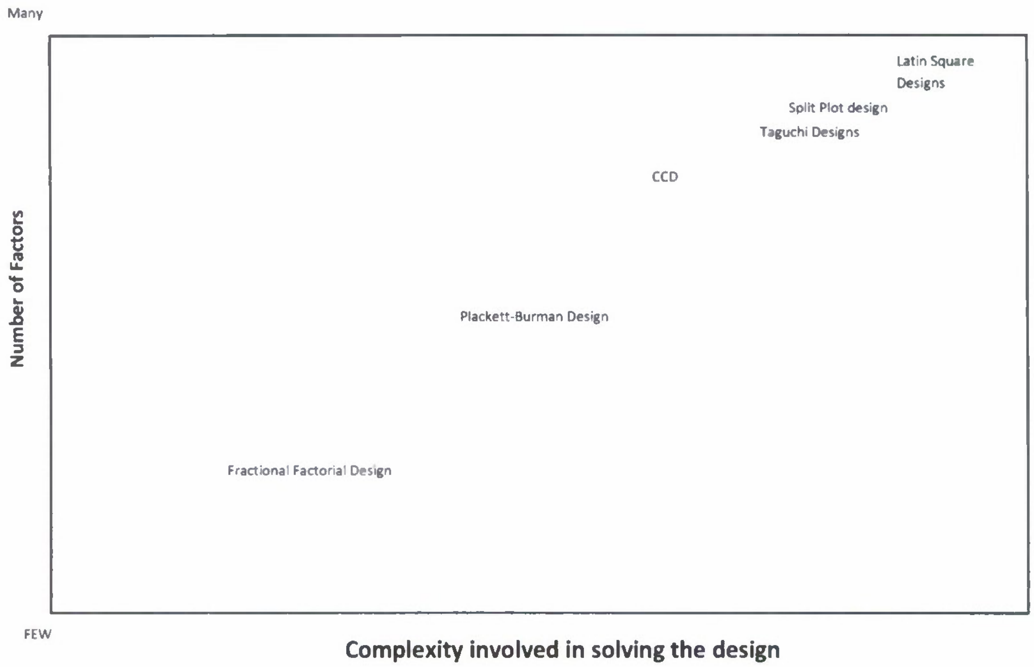


Figure 29. Complexity versus Number of factors.

This is the core of the thesis and thus a recommendation is given based on the designs and their behavior and features.

7.3 Comparison and Recommendation of DOE Softwares

A statistical analysis can be carried out only with a help of a software package. The easier and smaller experiments can be conducted intuitively and results calculated manually. But as the size of the design increases the analysis becomes more complex and cannot be performed efficiently without the help of the software [25]. The designs considered in this thesis solve large number of factors and are relatively difficult if not for these software packages. The packages today solve and analyze the designs with ease.

It is possible for them to carry out these by fitting data into inbuilt mathematical equations which predict the outcomes. These software packages help in coming up with optimal solutions which make it easy for engineers to carve up some savings in the industry. A comparative table is shown below (Table 16):

Table 16. Comparison of DOE softwares.

Name of the software	Features	Number of Factors	Cost
Design Ease	1. Breakthrough factors for process or product improvement. 2. Helps to set up and analyze general factorial, two-level factorial, fractional factorial and Plackett-Burman designs. 3. Numerical optimization	Up to 31 factors in fractional factorial and Plackett-Burman designs	For 1 – 2 $495.00 [26]
Design Expert	1. The peak of performance with the process or formulation. 2. Has features of Design Ease along with in-depth analysis of process factors or mixture components. 3. Offers rotatable 3D plots to help visualize the response surface. 4. Numerical optimization function present, which finds the most desirable factor settings for multiple responses simultaneously.	Include up to 256 runs and up to 8 blocks for 8 -15 factors	For 1 – 2 $995.00
ANOVA TM	1. Utilizes a complete set of orthogonal	Solves	For 1 it costs

	Arrays all the way to an L 108 or customize your own arrays. 2. Calculations for Dynamic Characteristic. 3. Individual Orthogonal Array Files and 4-way Analysis of Orthogonal Array or Single Factor. 4. Auto pooling on V, F and Rho%.	orthogonal Arrays up to L 108	Euro 995
DOE Kiss	1. An Excel Add In feature 2. Solves Taguchi, Plackett Burman design, Full and Fractional Dcsigns 3. Computer Aided Design Selection Wizard. 4. Custom Designs 5. Surface, Contour, and lntcraction Plots	It supports up to 26 factors and 1 response variable.	Price per copy is $249.00
Statistical sample planncr	1. Helps pick thc sample sizc. 2. Picks the best strategy that suits the experiment.		Pricc per copy $40.00
ECHIP	Handlcs all the aspects of a DOE.		Pricc per copy $1495.00
Minitab	1. Easy to use. 2. ANOVA, Regression analysis, Statistical analysis, Reliability analysis, multi variate analysis, design of cxperiment, rcsponsc surfacc, surface plots etc.		Price per copy $975.00
SAS	1. It has all the regular features of thc		Varies from

	statistical software along with it they have additional features like to solve MANOVA and Split Plot Design		vendor to vendor
SPSS	Has all features of SAS with compatibility with windows		For one copy $1600.00
Statistica	1. Comprehensive, user friendly interface, more than 11000 functions interactive 3D explorer to name a few		For one copy $1990.00

A detailed study of STASTISTICA and MINITAB has been done. Both these software packages have their own positive points and can be used to solve most of the large factor designs. STATISTICA is more robust software as it involves all the minute details involve theoretically in a design. But MINITAB is menu driven making it user friendly software used by most of the designers. Also it is more economical than the former.

8 Conclusions

A formalized, statistically-rigorous methodology has been proposed for the evaluation of any Data Fusion process. It is shown that the methodology requires the design of a separate Data Fusion process that specifically supports the T&E process by (a) providing a mathematical approach to the requirement to associate fused state estimates computed by the SUT prototype to truth states, and (b) providing the architecture and estimation processes for estimating the evaluation metrics of interest. Additionally, it is shown that the methodology also requires, due to the underlying stochastic nature of the DF process, integration of the methods of statistical experimental design and also, importantly, the associated methods of analysis that employ ANOVA techniques and other statistical analysis methods. Proof of concept experiments have been carried out to show representative application of the overall methods; these experiments and further elaborations of the methodological aspects are described.

The increased complexity of modern-day fusion-based tracking problems requires formalization and consistency in the PE process for fusion-developed estimates. This gets harder for Level 2, 3 and 4 fusion processes as explained in Section 2.3. We conducted empirical studies to bolster our understanding of the complex interdependencies in performance results from changing SUT/Scenario/PE parameters. This work suggests that the nature of the PE approach should build upon our familiarity with the "Fusion Tree" fusion process design for various applications. This work also shows that the quantitative effects of changing PE process techniques/parameters can significantly affect the MOP results.

In the second part of the research the need for large factor experiments was dealt by learning in depth about some methods that solved this problem. Some of these designs have real potential but have not been explored as they are not well known and they need a thorough perception. The mechanics behind these designs are hard and hence these are not very popular. A classification of these designs was carried out and recommendations were made after a thorough study. A comparative study based on the complexity, number of factors, the pros and cons has led to a subjective interpretation of the design which can be used as a guideline. Finally the software programs that aid in solving these designs were discussed and a comparison of these was made.

This research is a stepping stone in the world of huge intellectual opportunities to solve this non trivial problem. These topics need to be explored in detail and more specific guidelines need to be set. The designs can also be understood in a better way when the results could be quantified by plotting the number of runs needed by each design against the number of factors and also by conducting a trend analysis of the designs. Further an intensive study on the most suited software can be carried out and a manual developed which can aid experimenters to use these designs and encourages the use of the designs and exploit their potential.

Acknowledgement

This effort was motivated by the AFFTC at Edwards AFB and funded by AFOSR. We gratefully acknowledge the technical guidance from staff at AFFTC, and programmatic guidance from staff at AFOSR.

References

[1]. Sambhoos, K., Dastidar, S., Llinas, J., and Bowman, C., "Performance Evaluation Methods for Network Centric Data-Fusion Capable Tactical Platforms", Natl Symposium on Sensor and Data Fusion, Washington, DC, June 2006

[2]. Dastidar, S., Sambhoos, K., Bowman, C., and Llinas, J., "Performance Evaluation Methods for Data-Fusion Capable Tactical Platforms", Intl. Conf. on Information Fusion, Philadelphia, PA, July 2005

[3]. Llinas, J., Talreja, D., and Bowman, C., "Design of a Performance Evaluation Methodology for Data Fusion-based Multiple Target Tracking Systems: Part 2: Analysis Methodologies", Proceedings of Natl. Symp. On Sensor and Data Fusion (NSSDF) June 2004

[4]. Rawat, S., Llinas, J., and Bowman, C., "Design of a Performance Evaluation Methodology for Data Fusion-based Multiple Target Tracking Systems", presented at the SPIE Aerosense Conference, Orlando, FL, April 2003.

[5]. A. Hedayat and E. Seiden; "F-Square and Orthogonal F-Squares Design: A Generalization of Latin Square and Orthogonal Latin Squares Design"; The Annals of Mathematical Statistics, Vol. 41, No. 6 (Dec., 1970), pp. 2035-2044

[6]. Kramer, C.Y., Glass, S.;"Analysis of Variance of a Latin Square Design with Missing Observations"; Applied Statistics, Vol. 9, No. 1 (Mar., 1960), pp. 43-50

[7]. Gaito, J.; "The single latin square design in psychological research," Psychometrika, Springer, vol. 23(4),pages 369-378, December 1958.

[8]. Mandl, R.; "Orthogonal Latin squares: an application of experiment design to compiler testing"; Communications of the ACM, Volume 28 , Issue 10 (October 1985) Pages: 1054 - 1058

[9]. Tyssedal, J. and Samset, O.; "Analysis of the 12 Run Plackett-Burman Design", TECHREPORT 1997

[10]. Tyssedal, J.; Grinde, H. and Røstad, C.C.;"The Use of a 12-run Plackett-Burman Design in the Injection Moulding of a Technical Plastic Component", Quality and Reliability Engineering International Volume 22 Issue 6, Aug 2006, Pages 651 - 657

[11]. Devos, O., Fanget, B., Saber, A.-I., Paturel, L., Naffrechoux, E., and Jarosz, J.; "Use of a Plackett-Burman Design with Multivariate Calibration for the Analysis of Polycyclic Aromatic Hydrocarbons in Micellar Media by Synchronous Fluorescence"; Anal. Chem., 74, 3, 678 - 683, 2002.

[12]. Li J., Borror C.M., Cook C.A., Montogomery D.C; "Comparing prediction variance performance for variations of the Central Composite Design for 6 to 10 Factors"; Journal of Quality Technology, Quantitative Management, 2007.

[13]. Robert, C., Devillers, T., Wathelet, B., VanHerck, J.-C., and Paquot, M.; "Use of a Plackett-Burman Experimental Design to Examine the Impact of Extraction Parameters on Yields and Compositions of Pectins Extracted from Chicory Roots"; J. Agric. Food Chem., 54, 19, 7167 - 7174, 2006.

[14]. Vander Heyden, Y.; Dejaegher, B.; Capron, X.; "Fixing Effects and Adding Rows (FEAR) method to estimate factor effects in supersaturated designs constructed from Plackett-Burman designs"; Chemometrics and Intelligent Laboratory Systems, v85, n 2, 15 Feb. 2007, 220-31.

[15]. Vanderster, Dimopoulos, Sobie; "Improved Grid Metascheduler Design using the Plackett-Burman Methodology,"; pg. 9, 21st International Symposium on High Performance Computing Systems and Applications (HPCS'07), 2007.

[16]. Kowalski, S.M. and Potcner, K.J.; "How To Recognize A Split-Plot Experiment"; Quality Progress 1, November 2003, Pages - 60-66.

[17]. Sanchez, S.M. and Sanchez, P.J.; "Very large fractional factorial and central composite designs"; ACM Transactions on Modeling and Computer Simulation, v 15, n 4, Oct. 2005, 362-77.

[18]. Bjorkman, Colonel Eileen A. and Zeis Jr., Colonel Joseph E.; "Testing 2020: Evaluating weapons systems and capabilities in a joint environment"; Collection of Technical Papers - U.S. Air Force T and E Days, 2007, p 18-24.

[19]. Kothari. V.K., Sengupta. A.K. and Sensarma. J.K.; "Formation of Neps in Air-Jet Texturing Using a Central Composite Design"; Textile Research Journal, Vol. 66, No. 6, 384-388, 1996.

[20]. Kowalski,S.M.; Vining,G.G.; Montgomery.D.C.; Borror.C.M.; "Modifying a central composite design to model the process mean and variance when there are hard-to-change factors"; Journal of the Royal Statistical Society: Series C (Applied Statistics) Volume 55 Issue 5, 19 Oct 2006, Pages 615 - 630.

[21]. Robert Mee; "Noncentral Composite Designs"; Technometrics, V 43, No.1, Feb 2001.

[22]. T.N. Goh, "Quality and Reliability Engineering International," Vol. 9, 1993, pp 185-202.

[23]. Antony, J.; Perry, D.; Chengbo Wang; Kumar, M." An application of Taguchi method of experimental design for new product design and development process"; Assembly Automation, v 26, n 1, 2006, 18-24.

[24]. Moskowitz T.J., "An Analysis of Covariance Risk and Pricing Anomalies"; Review of Financial Studies, Volume 16, Number 2, 2003 , pp. 417-457(41).

[25]. Lee, J.J. and Kang, G., "An Educational Software for the Design of Experiments"; IASE/ISI Satellite, 2001.

[26]. "Software Price List" available via Link: http://www.statcase.com/swprice.html

[27]. Lloyd W. Condra, "Value-Added Management with Design of Experiments," Chapman & Hall, London, 1995.

[28]. Finney DJ., "Latin squares of the 6th order." Experientia, 2 (10), pp. 404-405 (1946).

[29]. Finney DJ., "Some orthogonal properties of the 4X4 and 6X6 Latin squares." Ann. Eugenic, 12 (4), pp. 213-219 (1945).

[30]. A. Hedayat and W. T. Federer, "On Embedding and Enumeration of Orthogonal Latin Squares," Ann. Math. Statist. Volume 42, Number 2 (1971), 509-516.

[31]. Freeman,G.H.; Some two-dimensional designs balanced for nearest neighbours. J.R.Statist.Soc., B41, 88-95, 1979.

[32]. Freeman,G.H.; Complete Latin squares and related experimental designs. J.R.Statist.Soc., B41, 253-262, 1979.

[33]. Freeman,G.H.; Further results on quasi-complete Latin squares. J.R.Statist.Soc., B43, 314-320, 1981.

[34]. Addleman, S. 1970. Variability of treatments and experimental units in the design and analysis of experiments. J. Amer. Stat. Asso. 65. Pg 1095-1108.

[35]. Armitage P. and Berry G., "Statistical Methods in Medical Research," 2nd edition., Blackwell Science, Oxford.

[36]. Clarke-O'Neill S., Pettersson L., Fader M., Dean G., Brooks R. & Cottenden A., "A multi-centre comparative evaluation: washable pants with an integral pad for light incontinence," Journal of Clinical Nursing, 11, 79-89, 2002.

[37]. Box,G. E. P. and Meyer,R. D., "Studies in Quality Improvement II. An Analysis for Unreplicated Fractional Factorials.," pages: 19, 1985.

[38]. Hamada, M. and Wu, C. F. J., "Analysis of designed experiments with complex aliasing," J. Quality Technology, 24, 130-137, 1992.

[39]. Hynen A., "Screening for Main and Interaction Effects with Plackett and Burman's 12 Run Design," Technical Report 1, Division of Quality Technology and Management, Linköping University, Sweden, 1996.

[40]. Box G.E.P. and Meyer R.D., "Finding the active factors in fractionated screening experiments," Journal of Quality Technology, 25(2), 94-105, 1993.

[41]. Yin Li; Zhiqiang Liu; Fengjie Cui; Zhisheng Liu; Hui Zhao; "Application of Plackett-Burman experimental design and Doehlert design to evaluate nutritional requirements for xylanase production by Alternaria mali ND-16", Applied microbiology and biotechnology, vol. 77, no2, pp. 285-291, 2007.

[42]. Box, G.E.P. and Jones, S., "Split-plot designs for robust product experimentation," J. Appl. Statist, 19, pp. 3-26, 1992.

[43]. Bisgaard, S., "The design and analysis of $2^{k-p}\times 2^{q-r}$ split-plot experiments," *J. Quality Technol,* 32, pp. 39–56, 2000.

[44]. Webb, D., Lucas, J. and Borkowski, J., "Factorial experiments when factor levels are not necessarily reset", Journal of Quality Technology, Vol. 36 No.1, pp.1-11, 2004.

[45]. Jones Bradley, Goos Peter., "A candidate-set-free algorithm for generating D-optimal split-plot designs," Journal of the Royal Statistical Society: series C: applied statistics, 56:3, p. 347-364, 2007.

[46]. Schoen, E. D., "Designing fractional two-level experiments with nested error structures," J. Appl. Statist., 26, 495-508, 1999.

[47]. Box, G.E.P. and Wilson, K.B., "Experimental attainment of optimum conditions," Journal of the Royal Statistical Society, 13, pp. 1-45, 1951.

[48]. Hader, R.J. and Park, S.H., "Slope-rotatable central composite designs", Technometrics, 20, 413-417, 1978.

[49]. Bassem R. Mahafza and Atef Elsherbeni, "MATLAB Simulations for Radar Systems Design," CRC Press, December 2003.

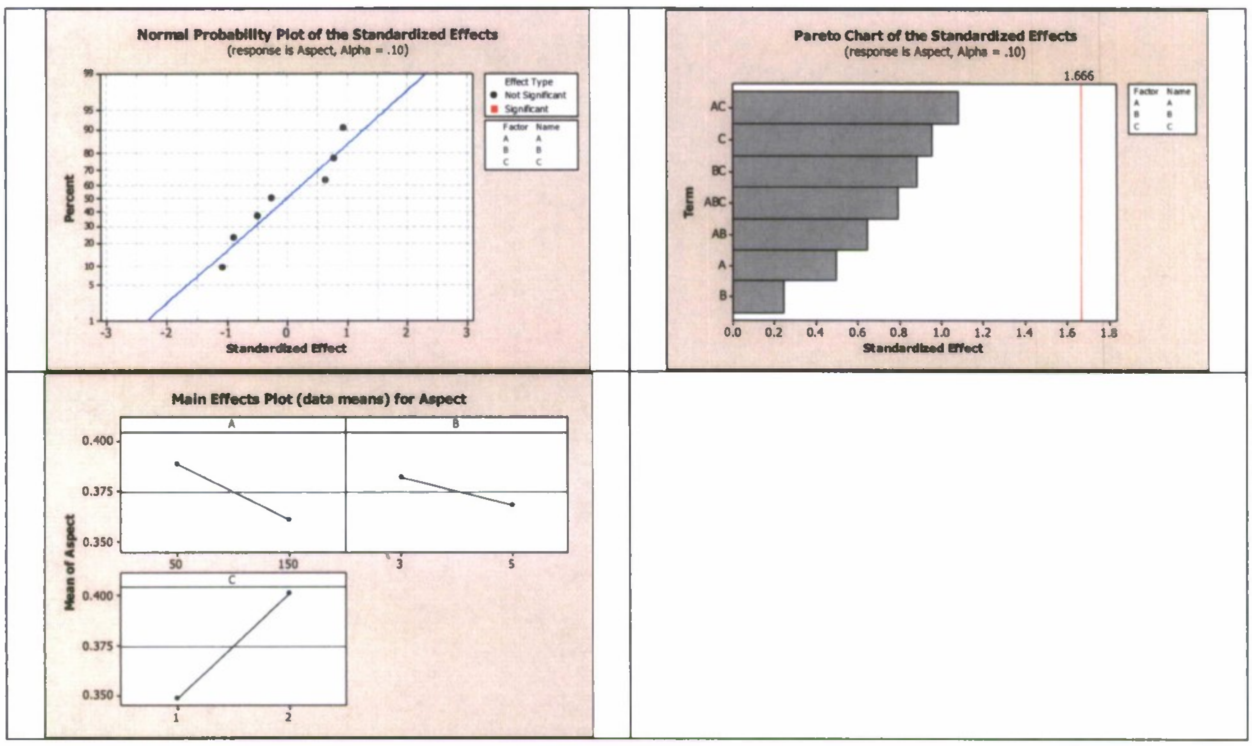
Appendix

A. L2 and L3 Tier 0 DOE Charts

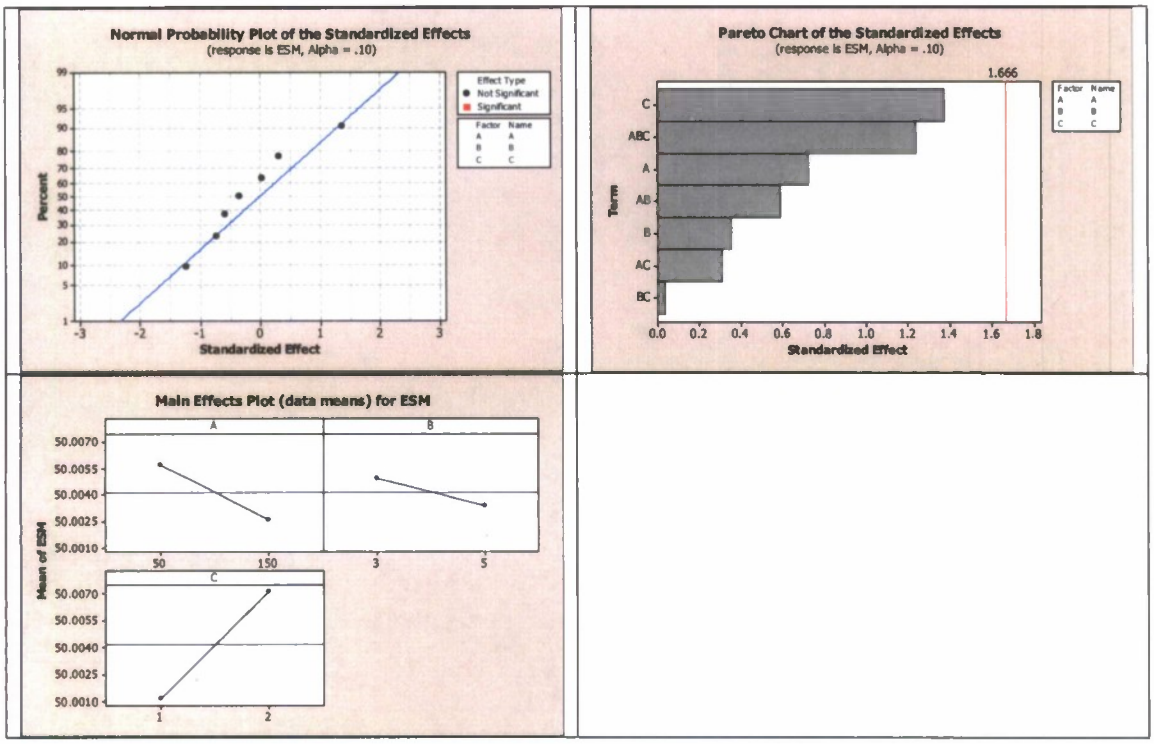
This section provides the Tier 0 DOE charts conducted in Section 5. The three factors SUT Design Gating Factor, PE Gating Factor and PE Design at two levels each are tested to find which of these factors affect the MOPs significantly. In Tier 0 we have three sensors on 2 platforms and they do not fuse any data within or across platform. Hence we have to only analyze track-to-truth associations for each of the MOPs. The summary of the results is shown in Table 3. Here for each MOP we have the Normal Probability plot and Pareto chart which summarizes the significant factors. Then for the significant factors we plot the main effects plot which tells us how the change in factor affects the MOP. For the significant interactions we plot the interaction plot which shows the effect of change in factor level combination on MOP.

After taking a look at the summary (Table 3), we can say that SUT Design Gating Factor is comparatively more significant than PE Gating Factor and PE Design. SUT Design Gating Factor appears to be a significant factor in nearly all the Tier 0 DOE runs. So at Tier 0 we must be sensitive towards selection of SUT Design Gating Factor.

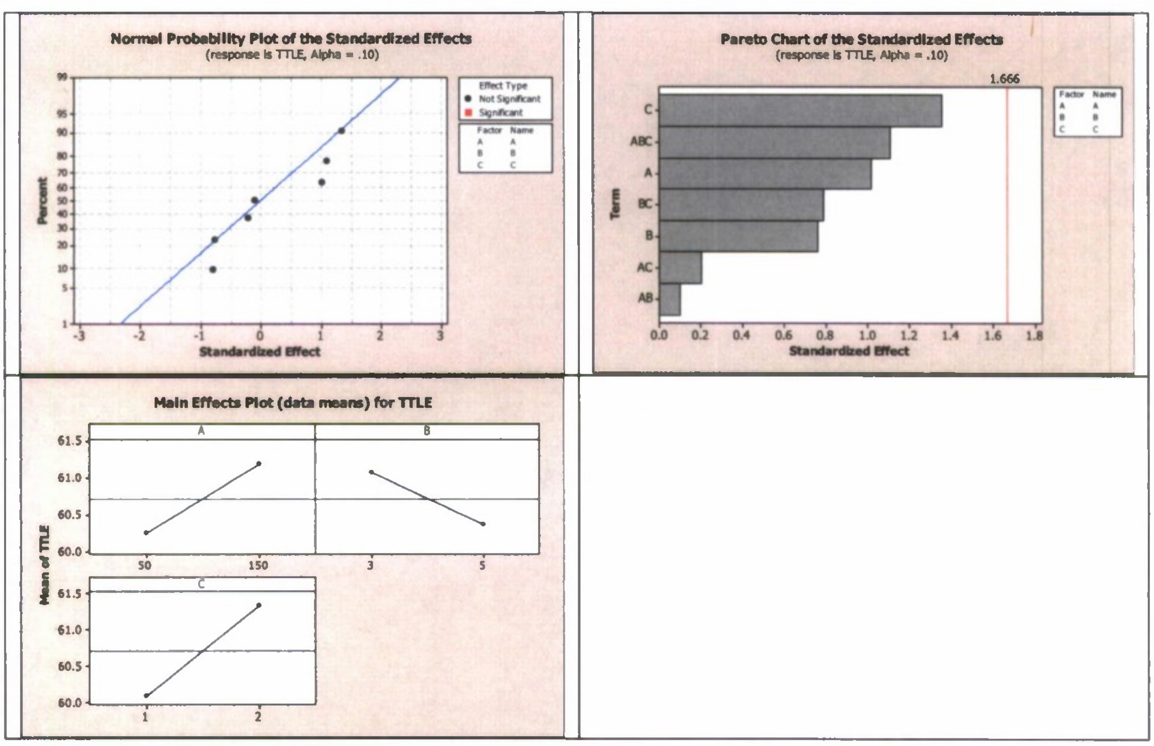
Radar Track 1 to Truth Aspect Consistency:



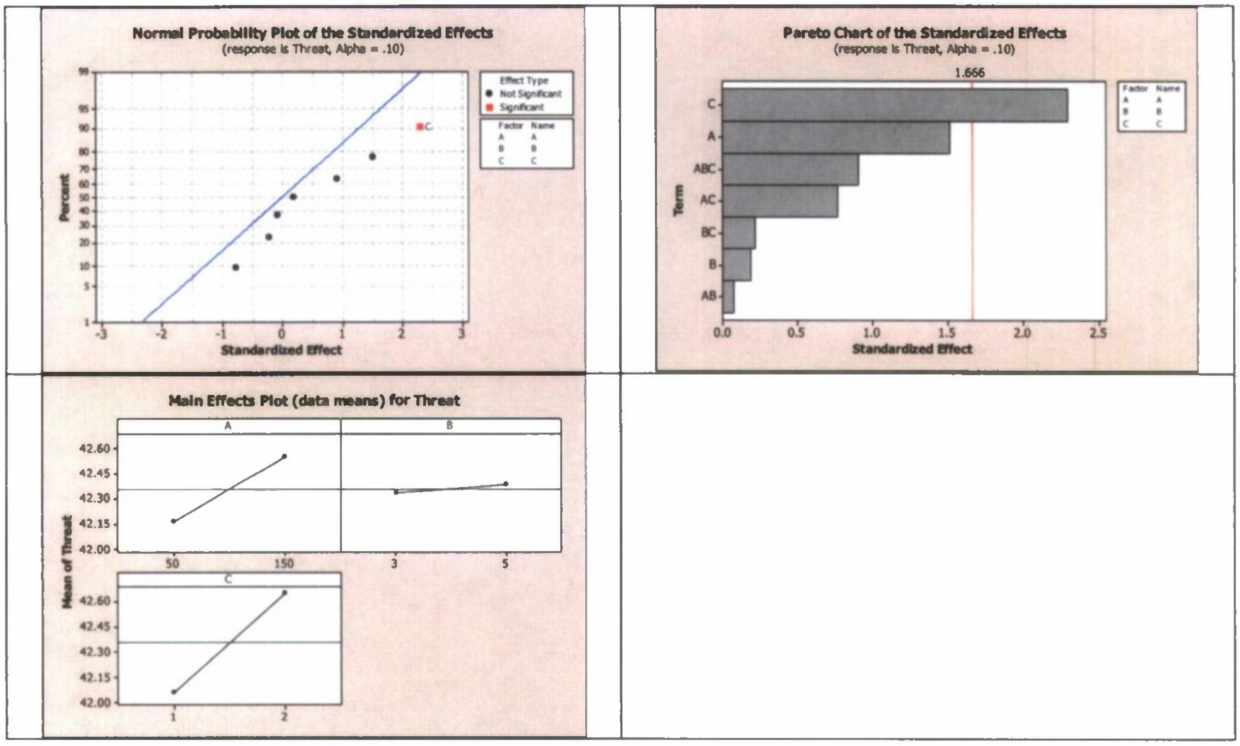
Radar Track 1 to Truth ESM Consistency:



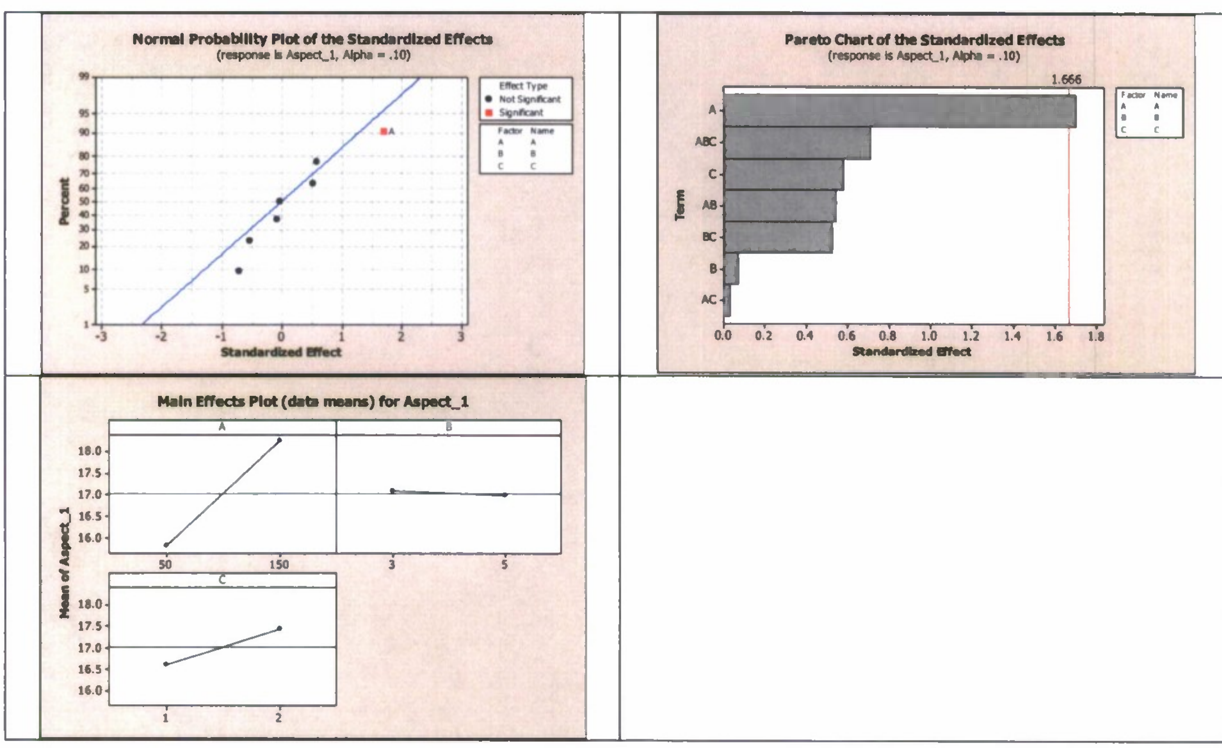
Radar Track 1 to Truth TTLE Consistency:



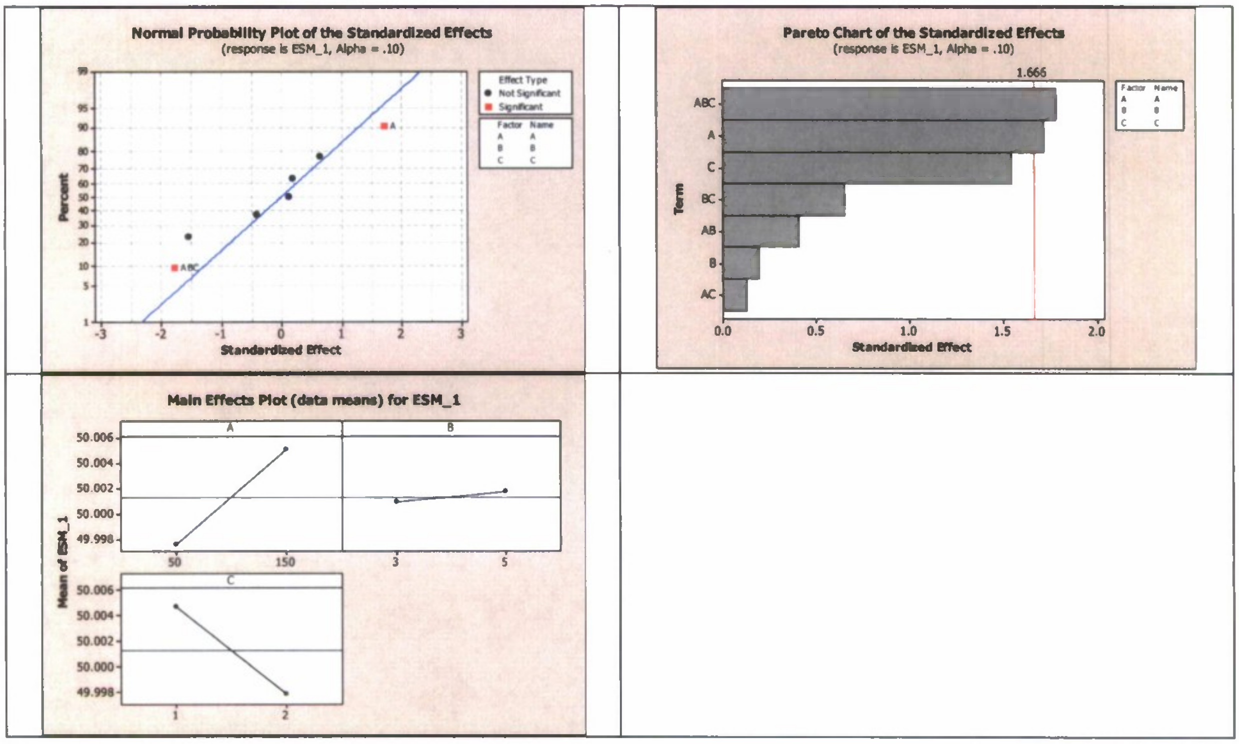
Radar Track 1 to Truth Threat Consistency:



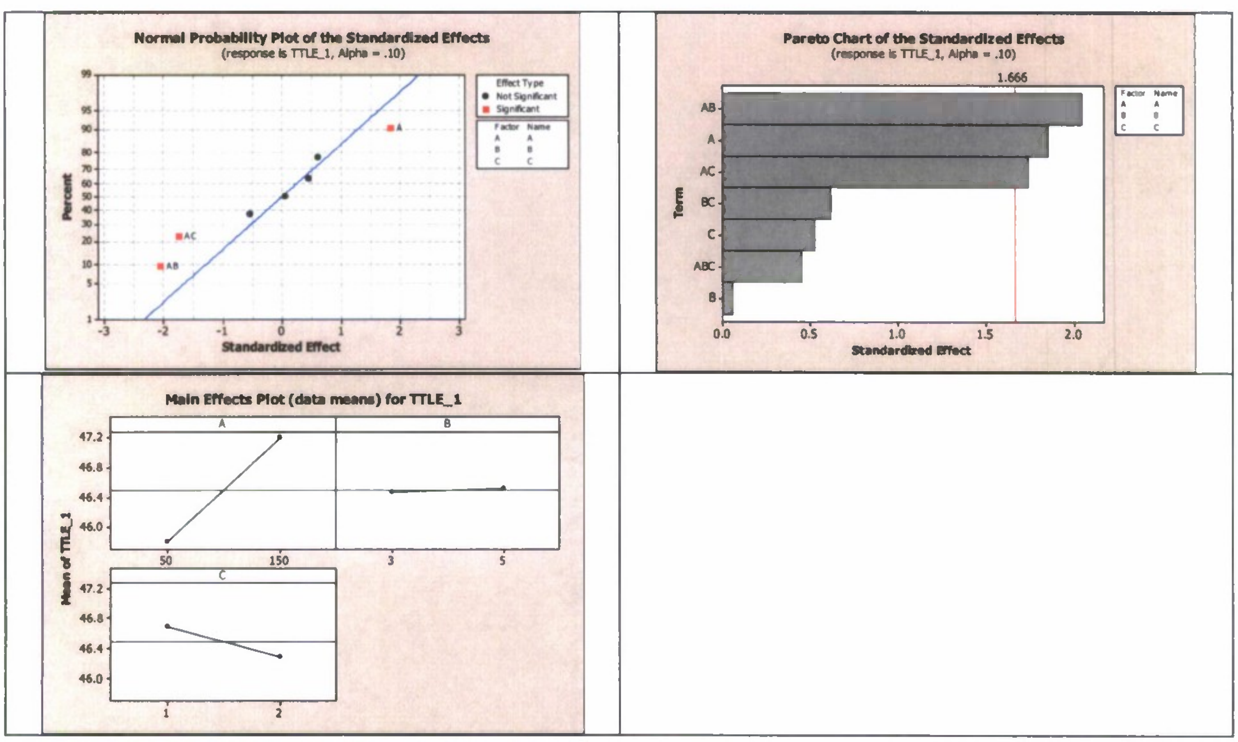
Radar Track 2 to Truth Aspect Consistency:



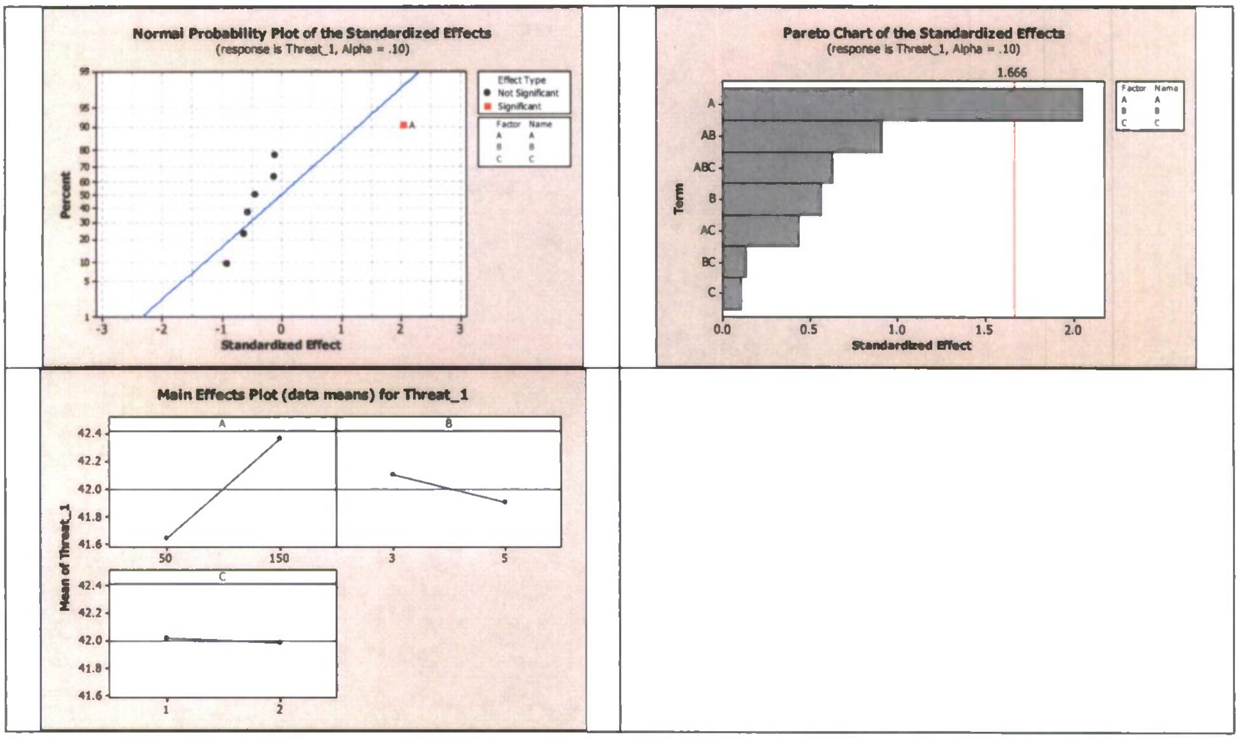
Radar Track 2 to Truth ESM Consistency:



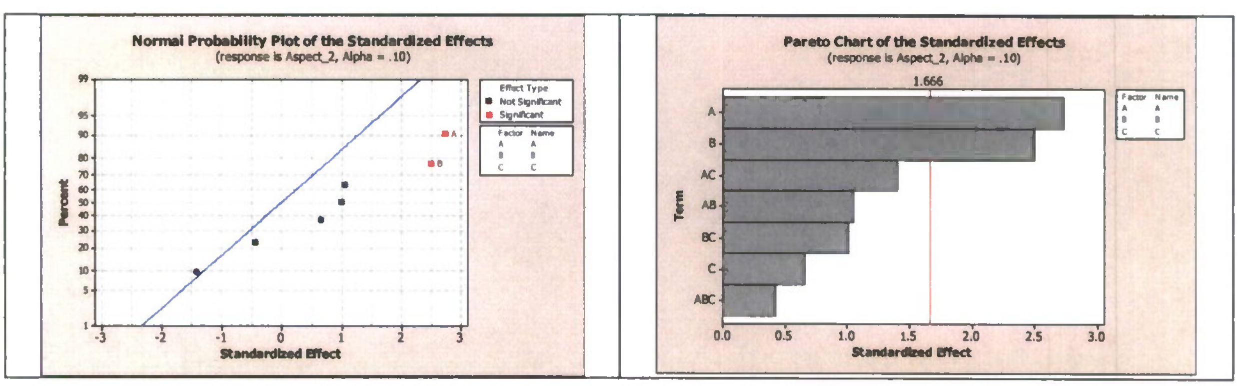
Radar Track 2 to Truth TTLE Consistency:

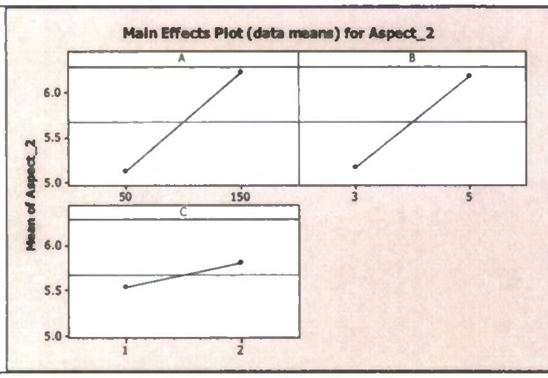


Radar Track 2 to Truth Threat Consistency:

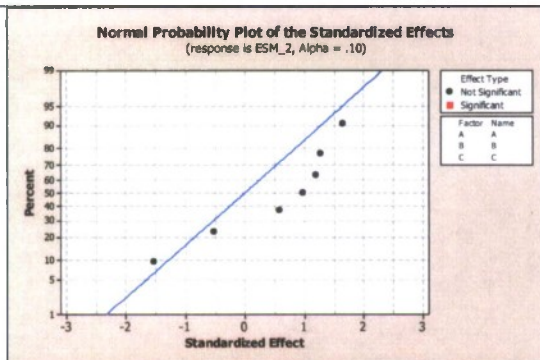


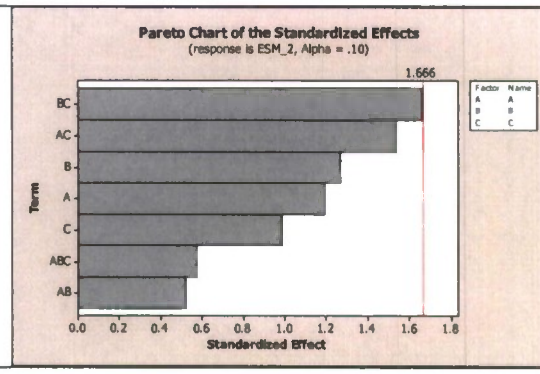
ESM Track 1 to Truth Aspect Consistency:

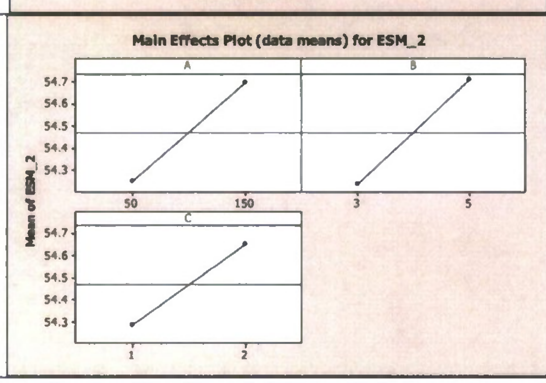




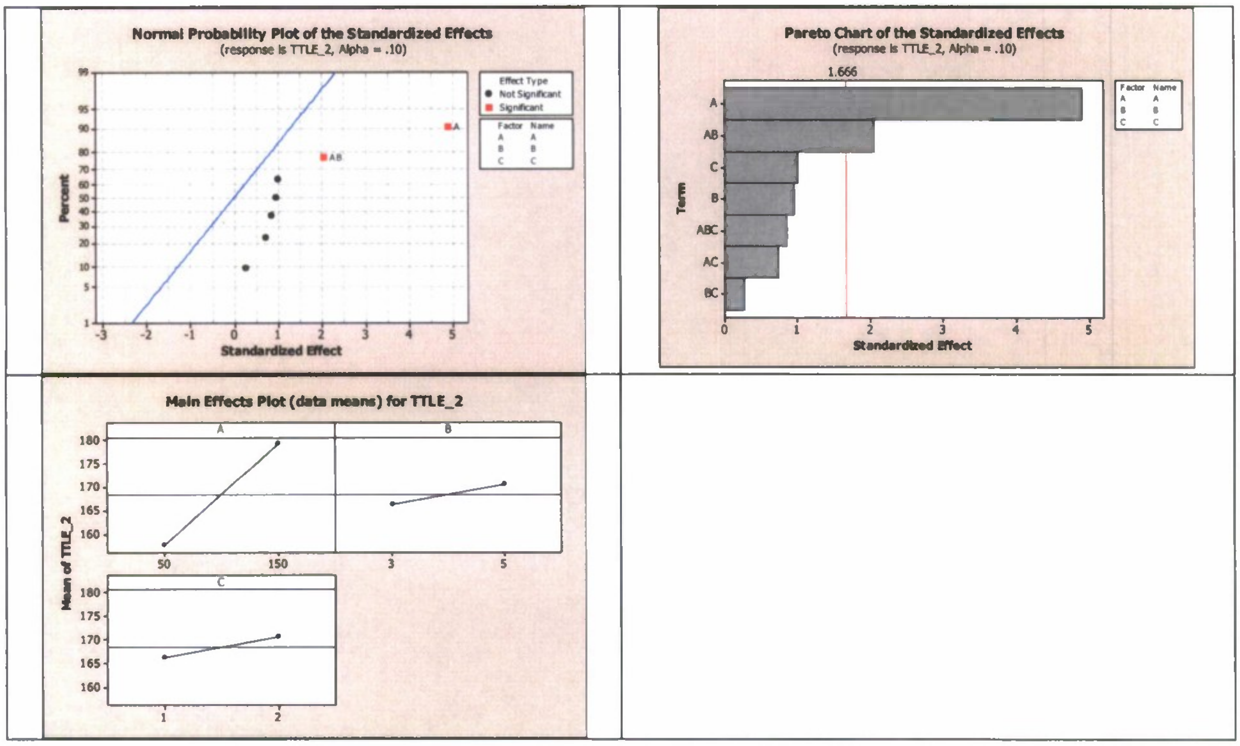
ESM Track 1 to Truth ESM Consistency:



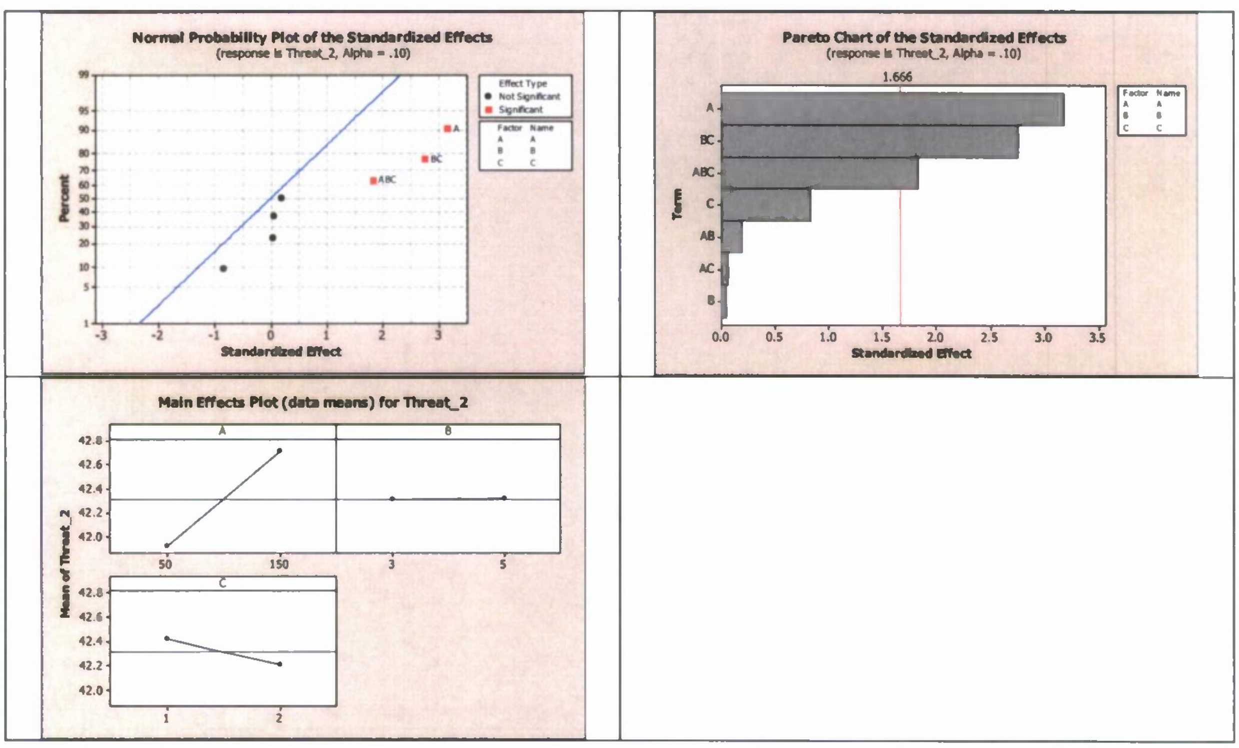




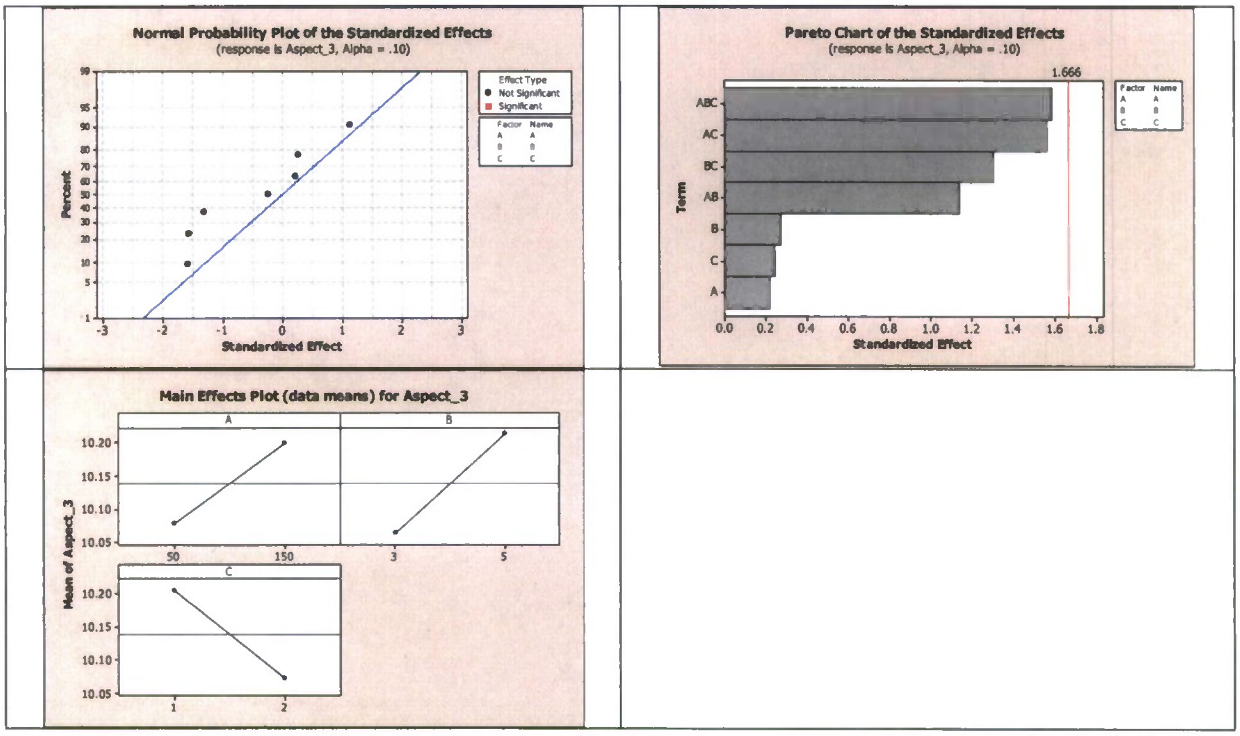
ESM Track 1 to Truth TTLE Consistency:



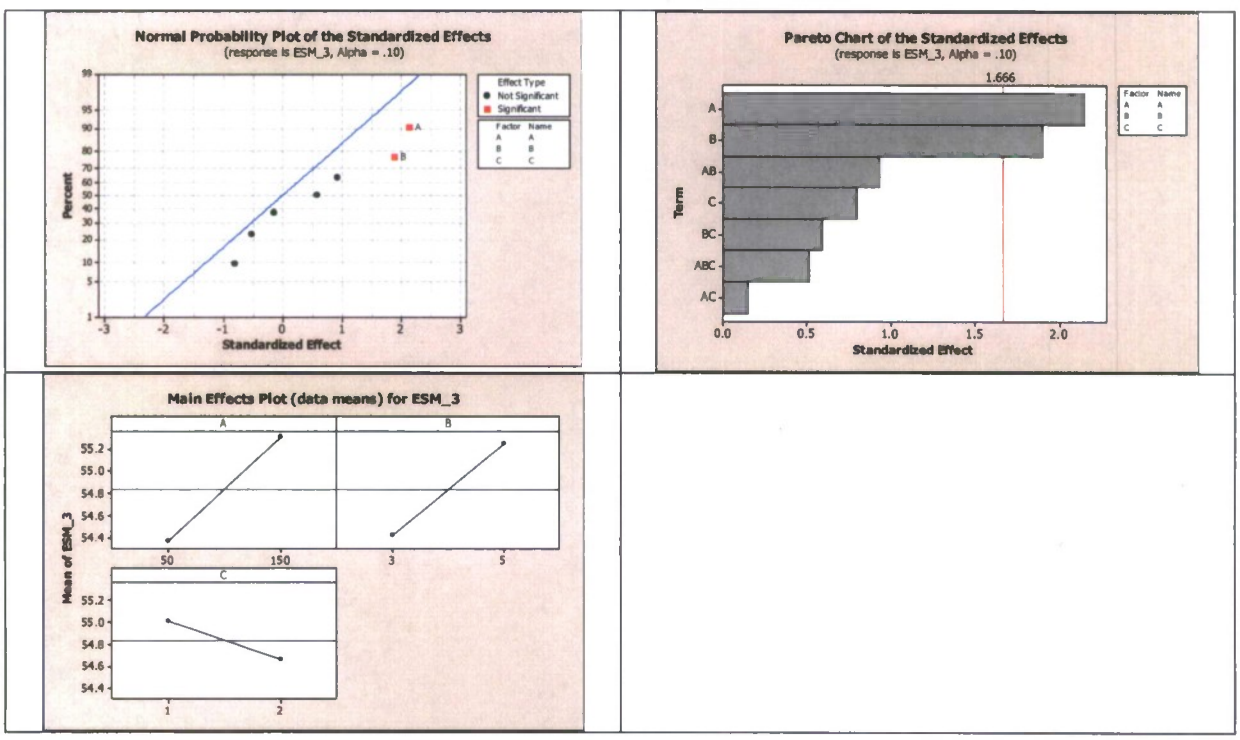
ESM Track 1 to Truth Threat Consistency:



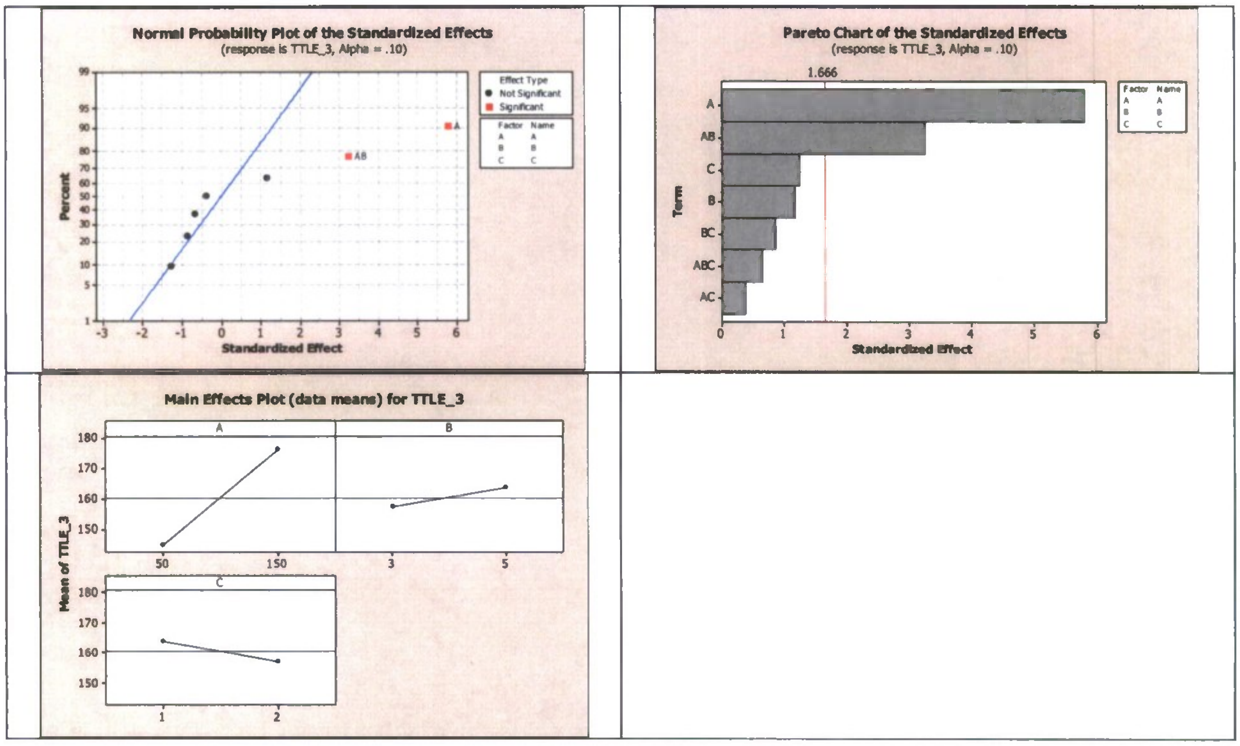
ESM Track 2 to Truth Aspect Consistency:



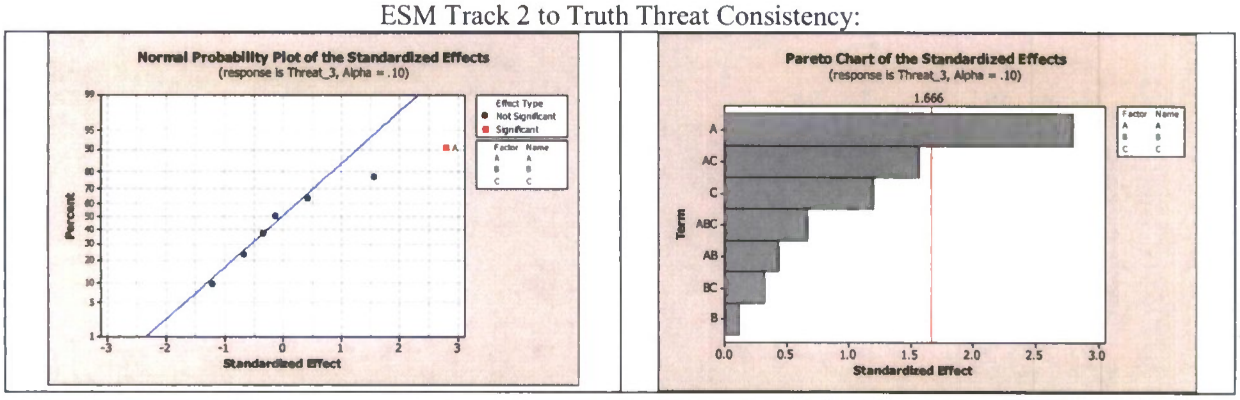
ESM Track 2 to Truth ESM Consistency:

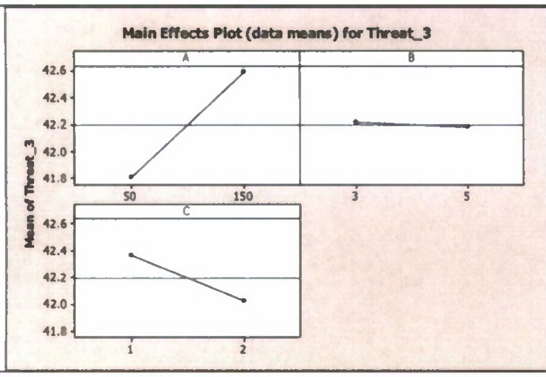


ESM Track 2 to Truth TTLE Consistency:

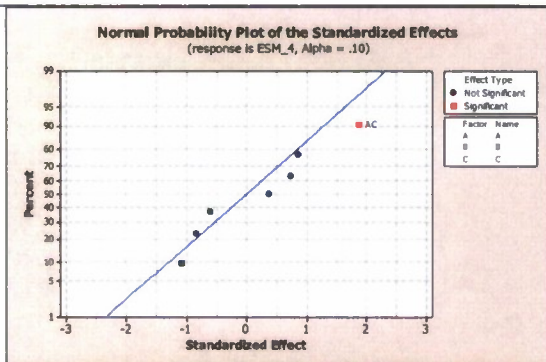


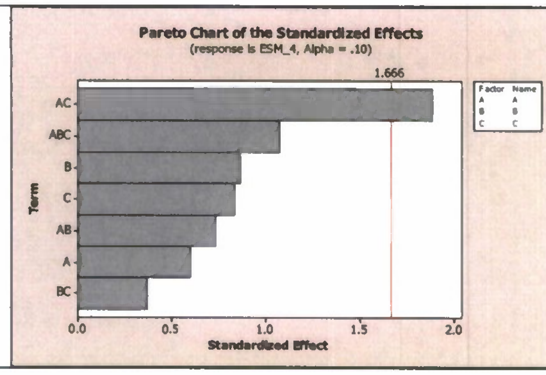
ESM Track 2 to Truth Threat Consistency:

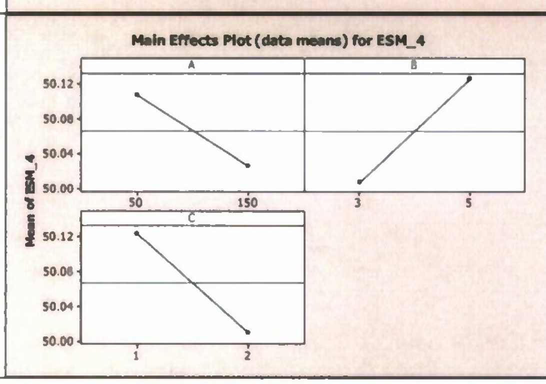




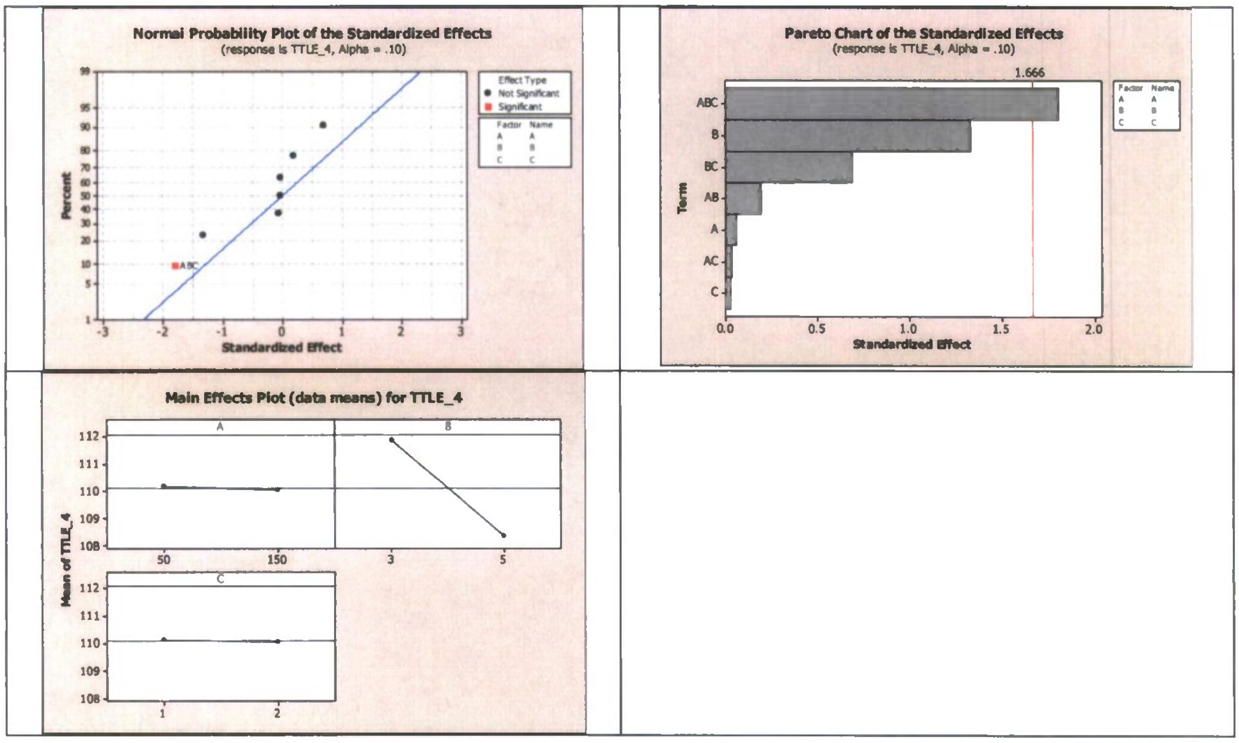
IRST Track 1 to Truth ESM Consistency:



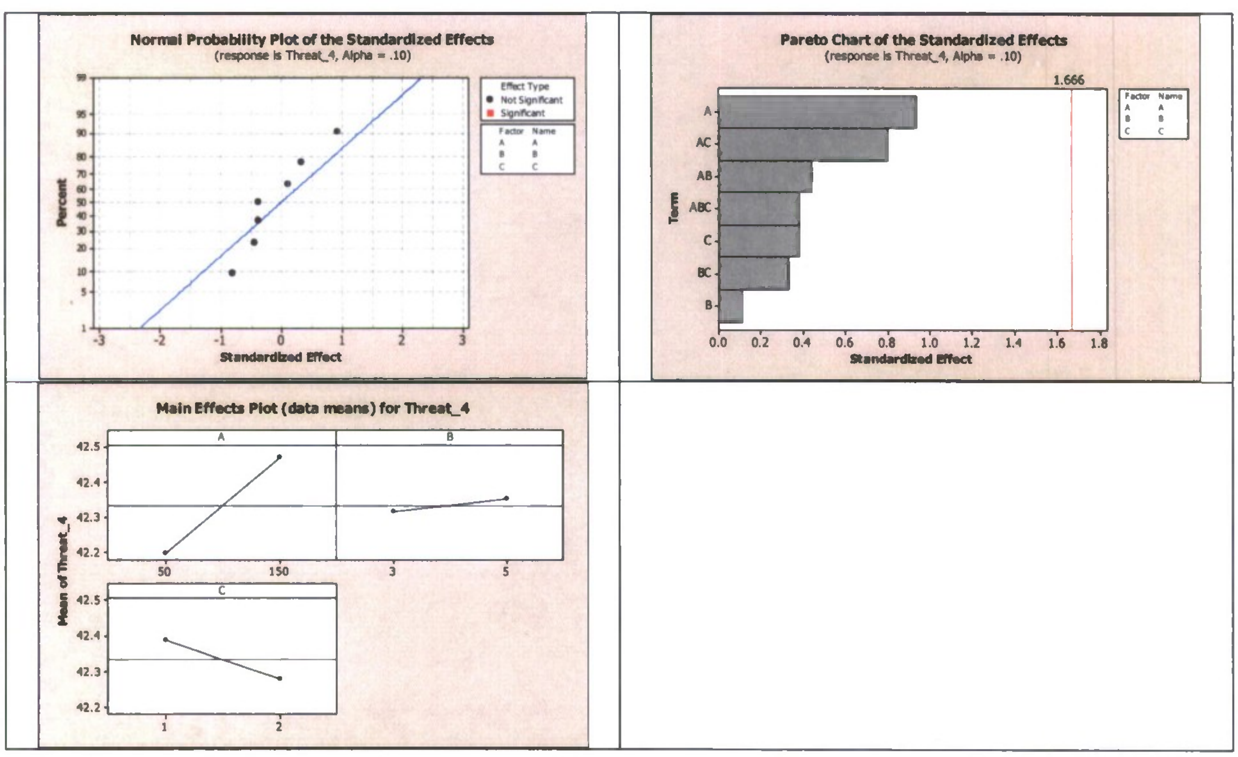




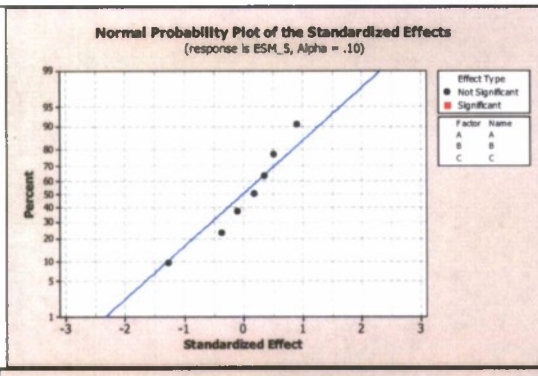
IRST Track 1 to Truth TTLE Consistency:

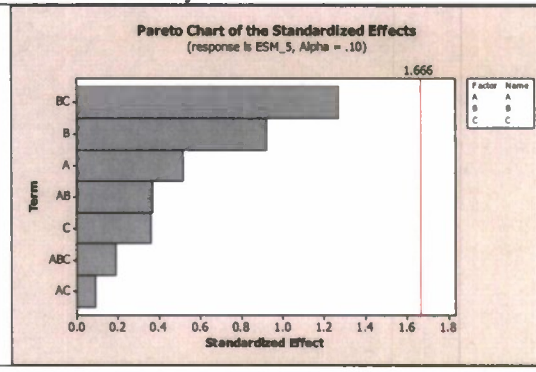


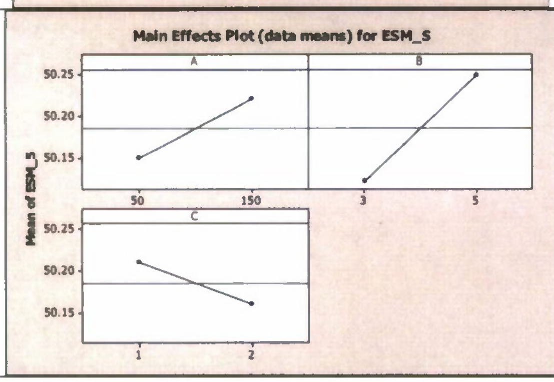
IRST Track 1 to Truth Threat Consistency:



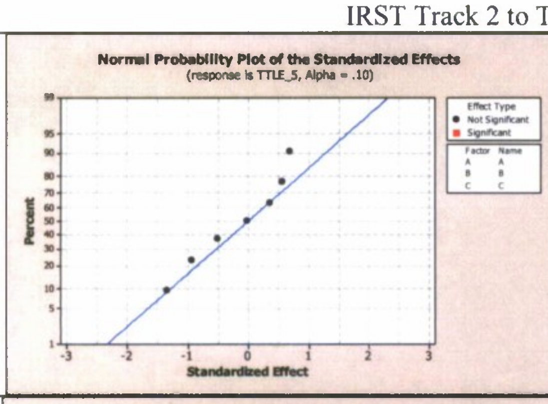
IRST Track 2 to Truth ESM Consistency:

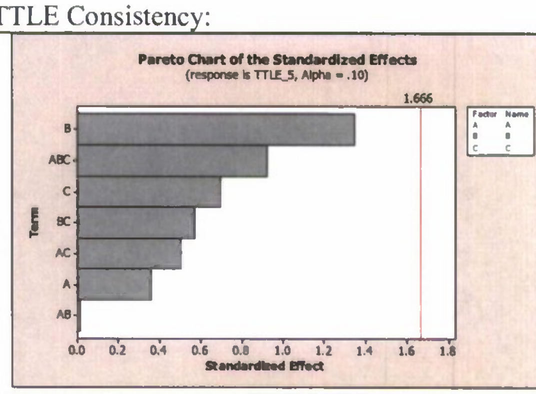


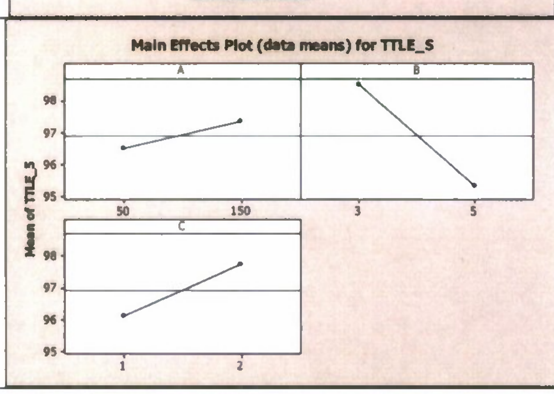




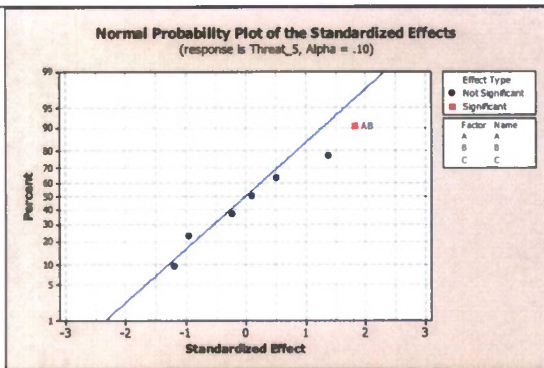
IRST Track 2 to Truth TTLE Consistency:

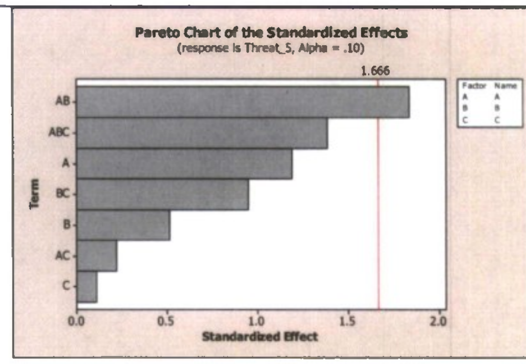


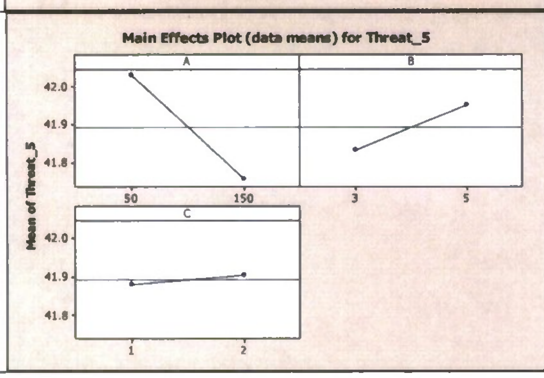




IRST Track 2 to Truth Threat Consistency:





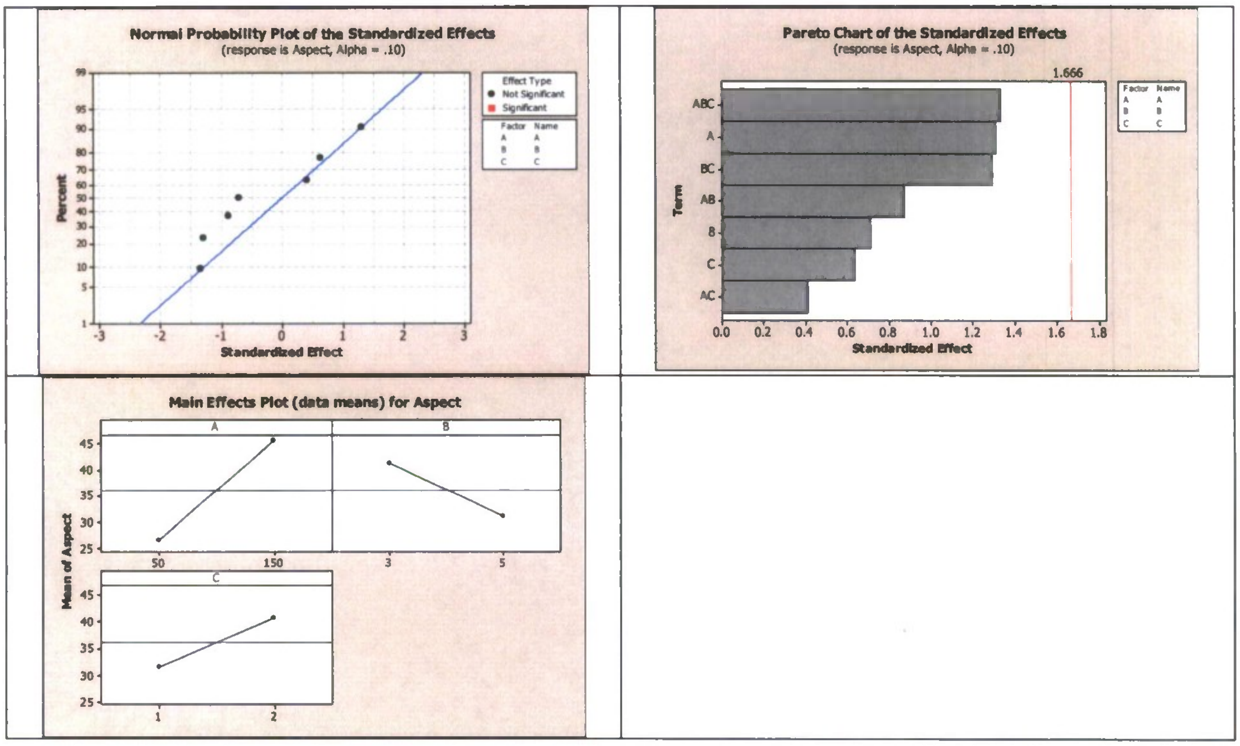


B. L2 and L3 Tier 1 DOE Charts

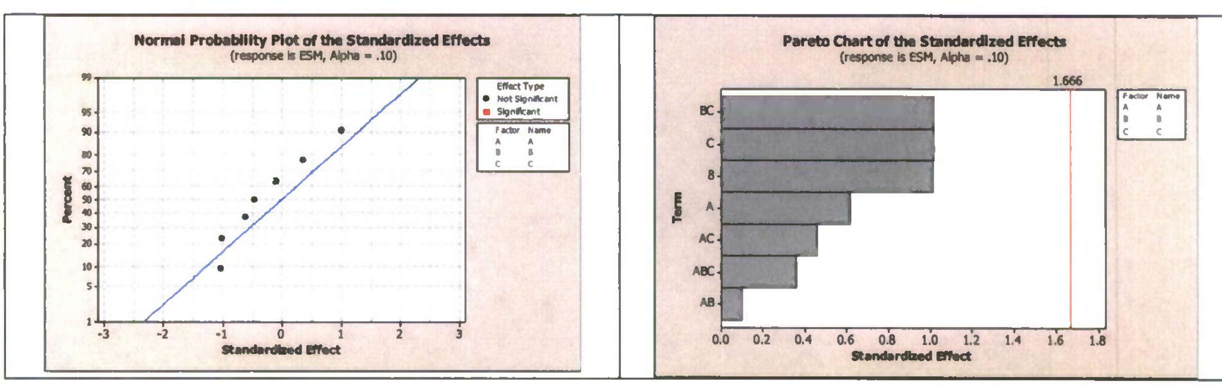
This section provides the Tier 1 DOE charts conducted in Section 5. The three factors SUT Design Gating Factor, PE Gating Factor and PE Design at two levels each are tested to find which of these factors affect the MOPs significantly In Tier 1 we have three sensors on 2 platforms and they fuse data within platform (not across platform). So we have to analyze track-to-truth and track-to-track associations for each of the MOPs. The summary of the results is shown in Table 4. Here for each MOP we have the Normal Probability plot and Pareto chart which summarizes the significant factors. Then for the significant factors we plot the main effects plot which tells us how the change in factor affects the MOP. For the significant interactions we plot the interaction plot which shows the effect of change in factor level combination on MOP.

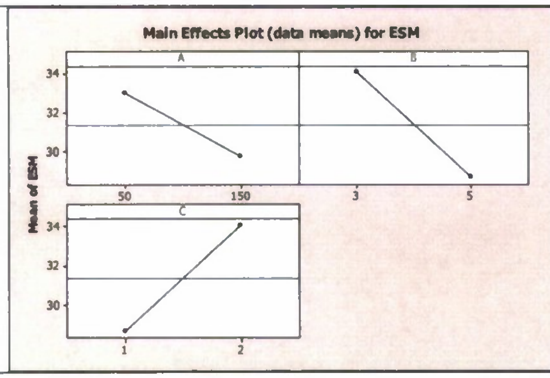
After taking a look at the summary Table 4, we can say that PE Design is comparatively more significant than PE Gating Factor and SUT Design Gating Factor. PE Design appears to be a significant factor in nearly all the Tier 1 DOE runs. So at Tier 1 we must be sensitive towards selection of PE Design

Track To Track Aspect Consistency:

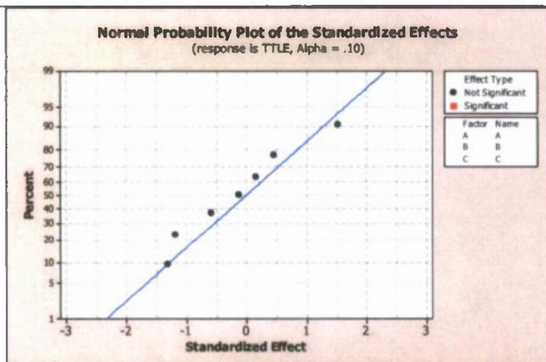


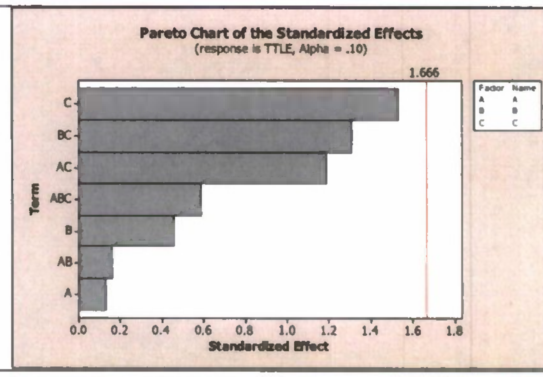
Track To Track ESM Consistency:

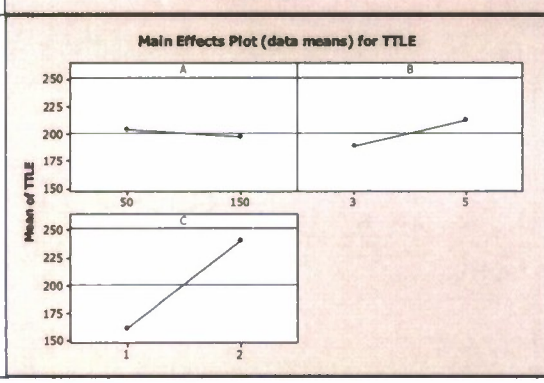




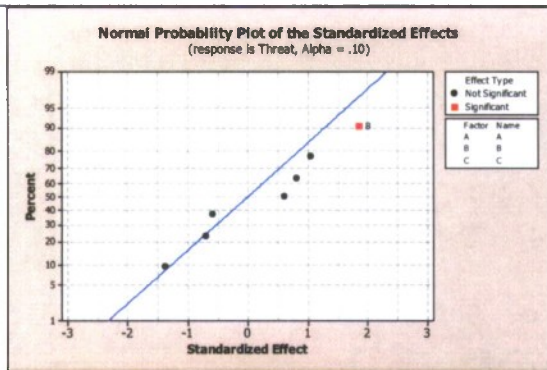
Track To Track TTLE Consistency:

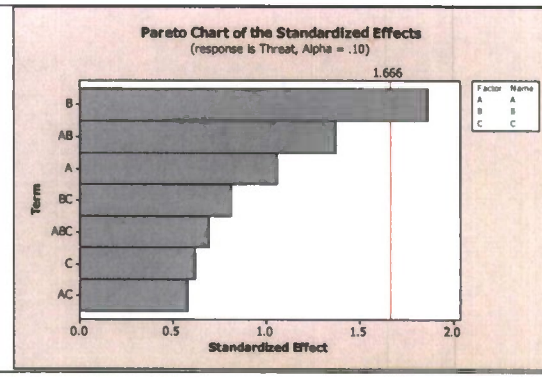


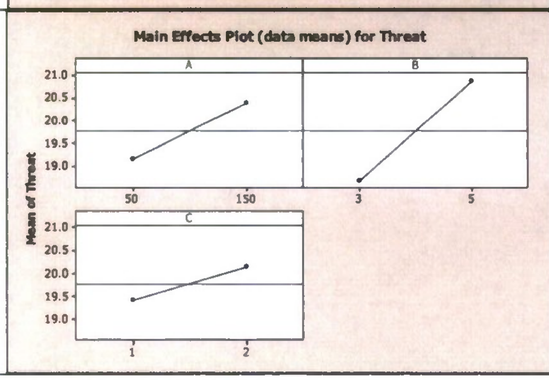




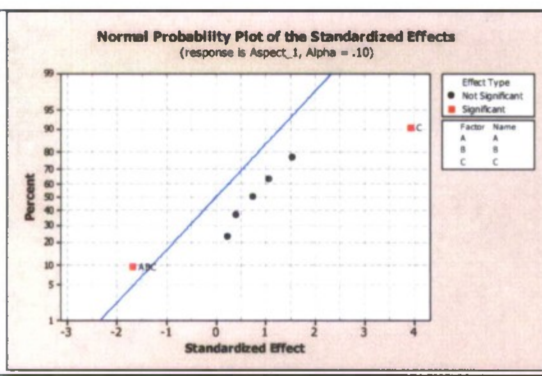
Track To Track Threat Consistency:

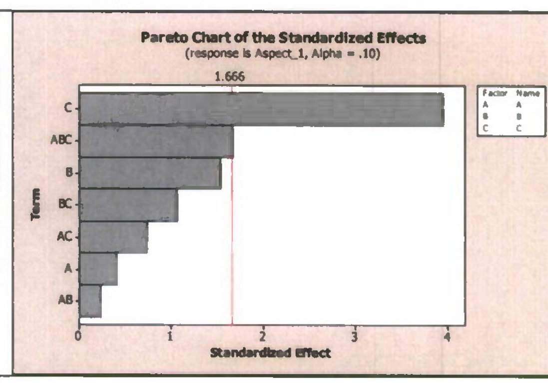


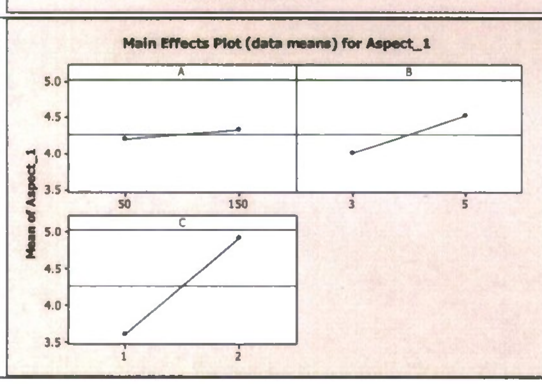




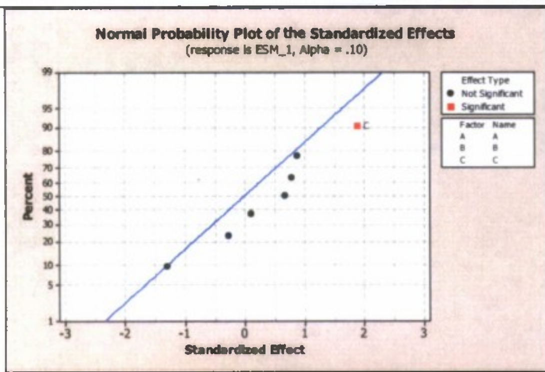
Track 1 to Truth Aspect Consistency:

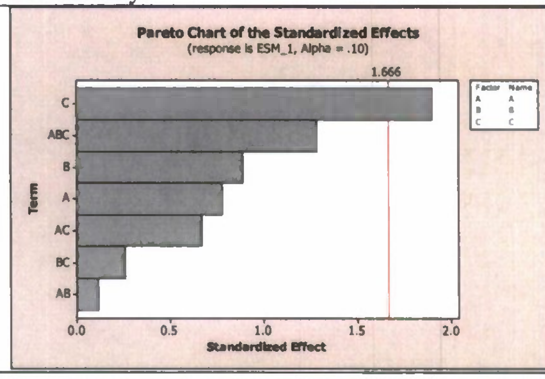


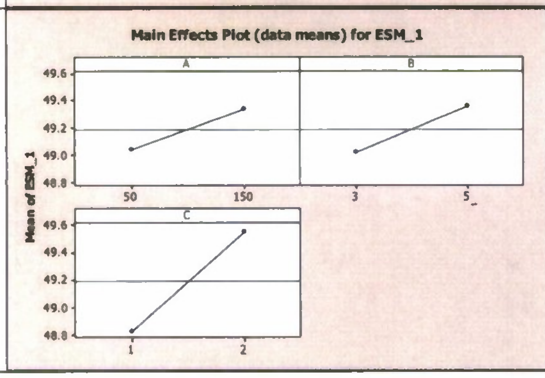




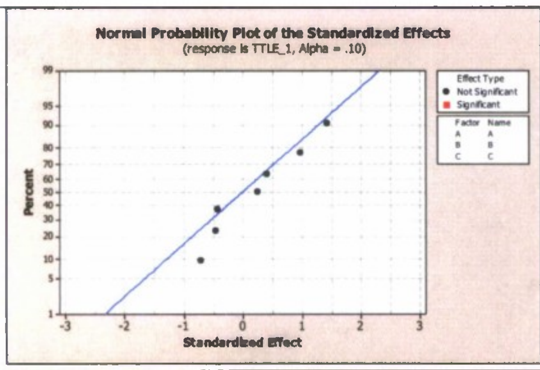
Track 1 to Truth ESM Consistency:

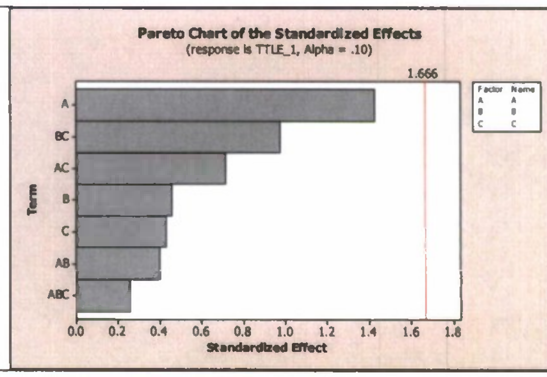


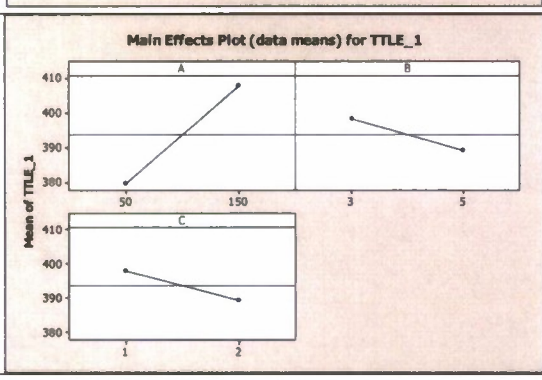




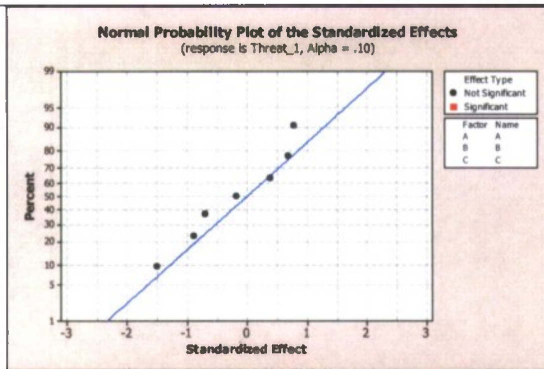
Track 1 to Truth TTLE Consistency:

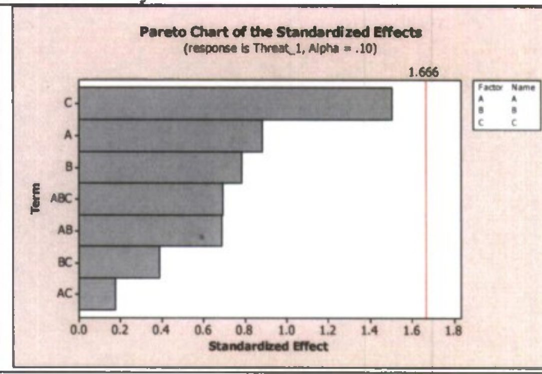


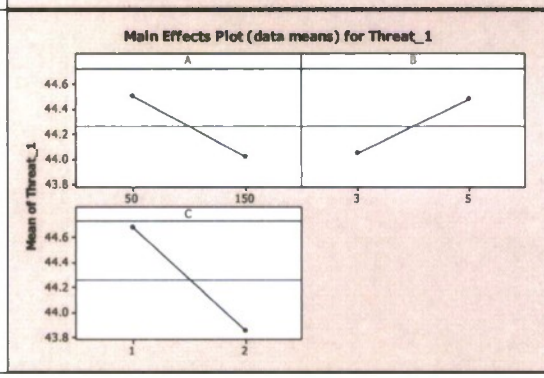




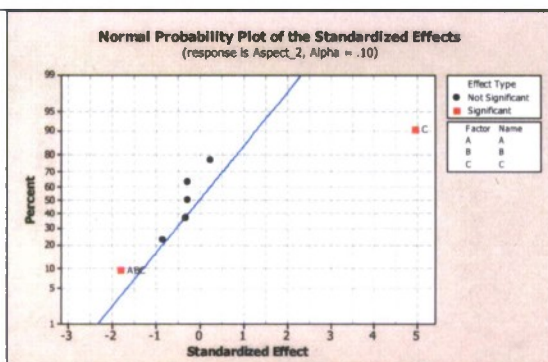
Track 1 to Truth Threat Consistency:

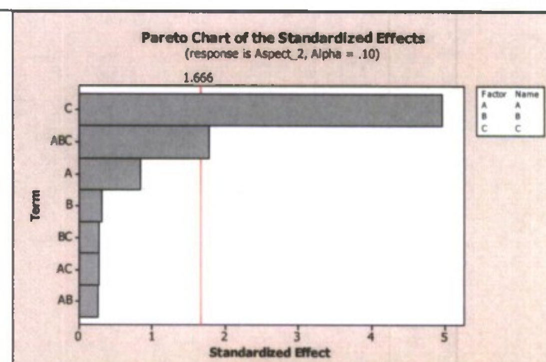


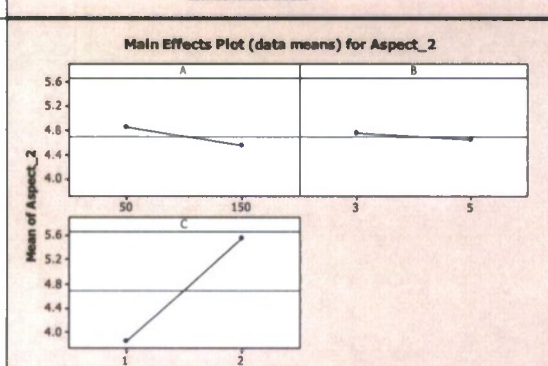




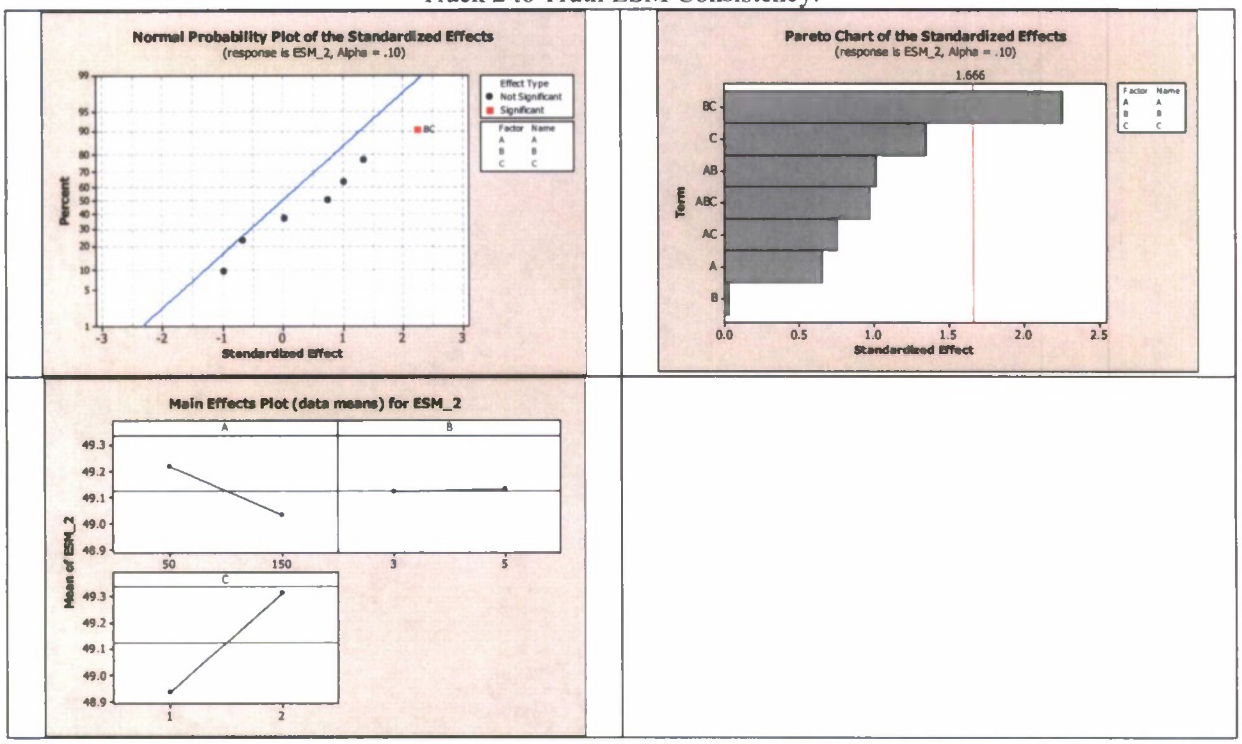
Track 2 to Truth Aspect Consistency:



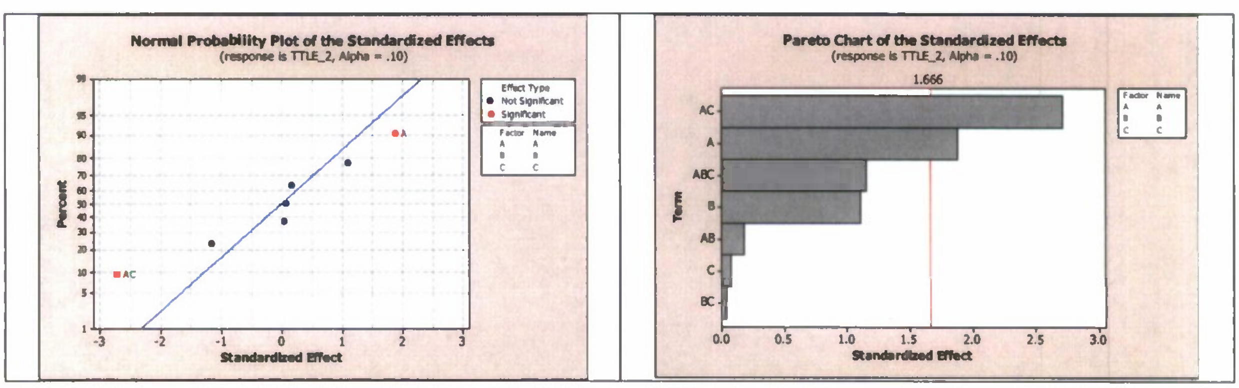


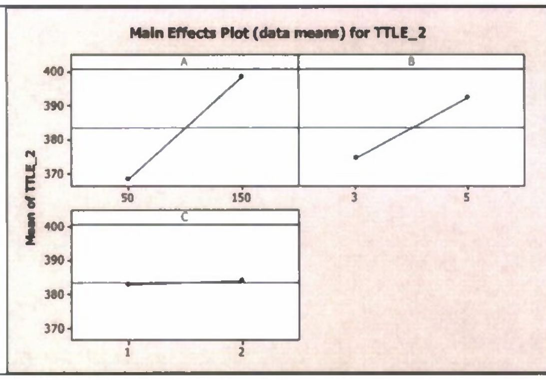


Track 2 to Truth ESM Consistency:

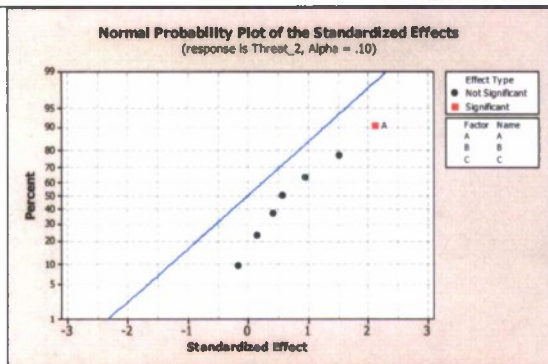


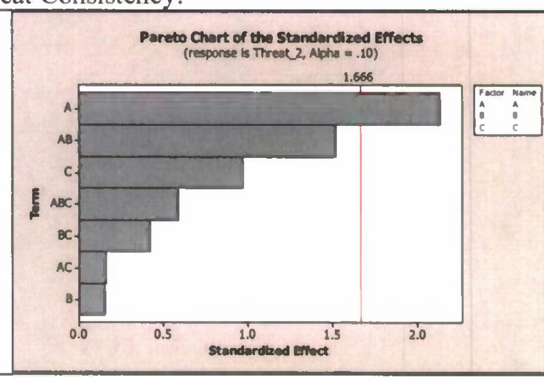
Track 2 to Truth TTLE Consistency:

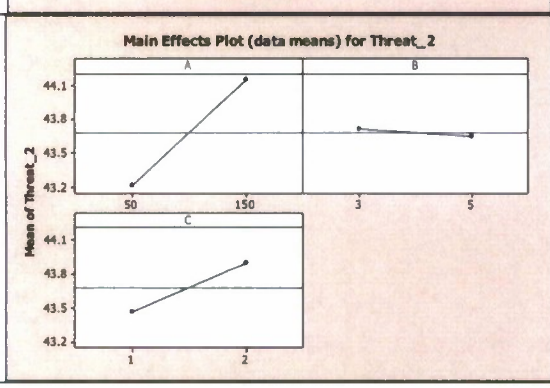




Track 2 to Truth Threat Consistency:







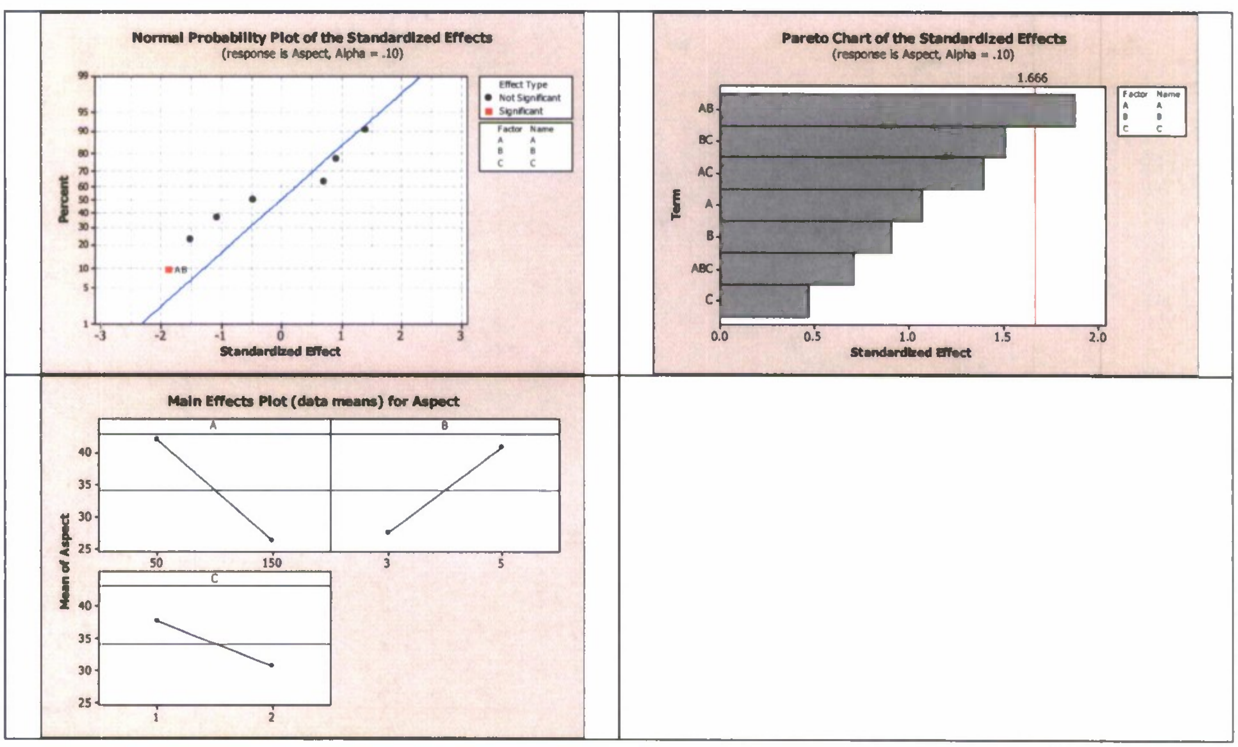
C. L2 and L3 Tier 2 DOE Charts

This section provides the Tier 2 DOE charts conducted in Section 5. The three factors SUT Design Gating Factor, PE Gating Factor and PE Design at two levels each are tested to find which of these factors affect the MOPs significantly. In Tier 2 we have three sensors on 2 platforms and they fuse data within and across platforms. So we have to analyze track-to-truth and track-to-track associations for each of the

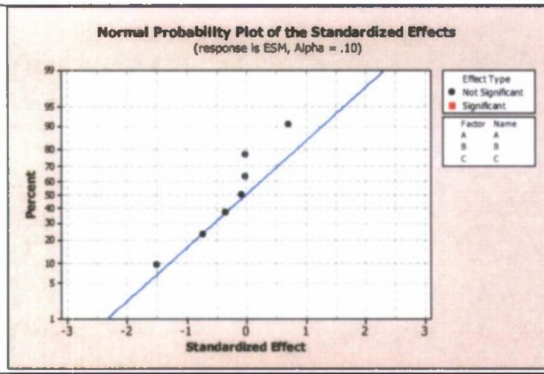
MOPs. The summary of the results is shown in Table 5. Here for each MOP we have the Normal Probability plot which summarizes the significant factors. Then for the significant factors we plot the main effects plot which tells us how the change in factor affects the MOP. For the significant interactions we plot the interaction plot which shows the effect of change in factor level combination on MOP.

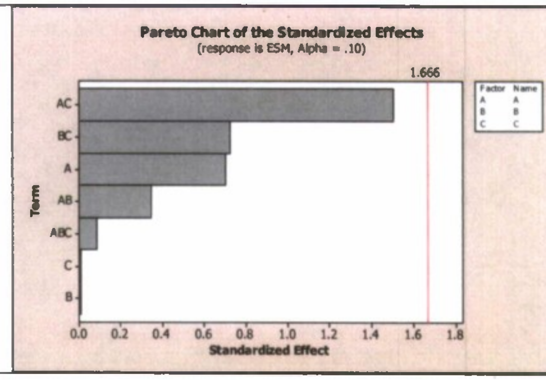
After taking a look at the summary Table 5, we can say that PE Design is comparatively more significant than PE Gating Factor and SUT Design Gating Factor. In this case some of the two way interactions involving PE Design are significant which suggests that fusing data across platforms reduces the discrepancies in the input data.

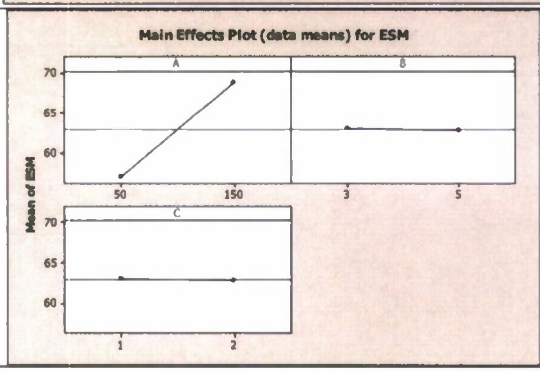
Track to Track Aspect Consistency:



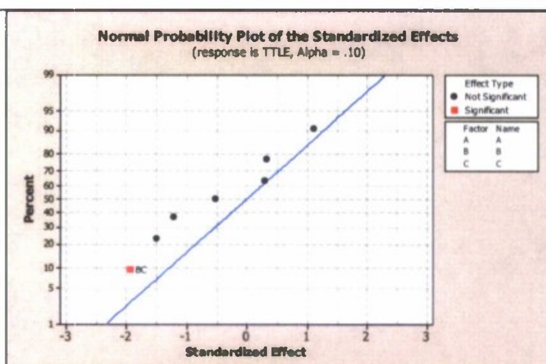
Track to Track ESM Consistency:

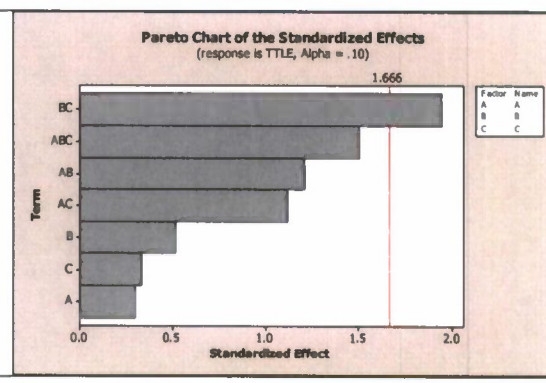


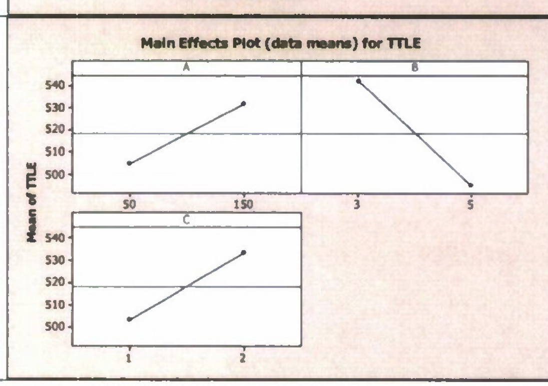




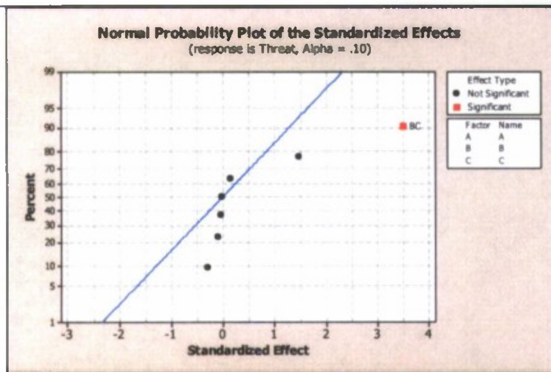
Track to Track TTLE Consistency:

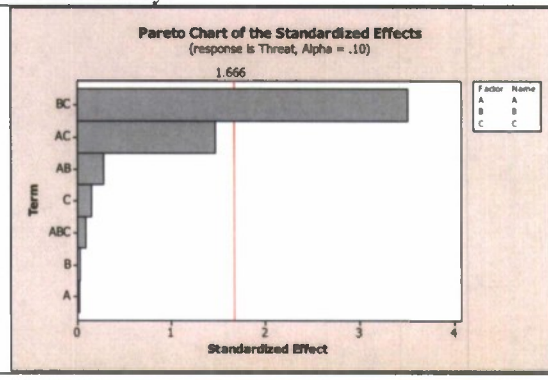


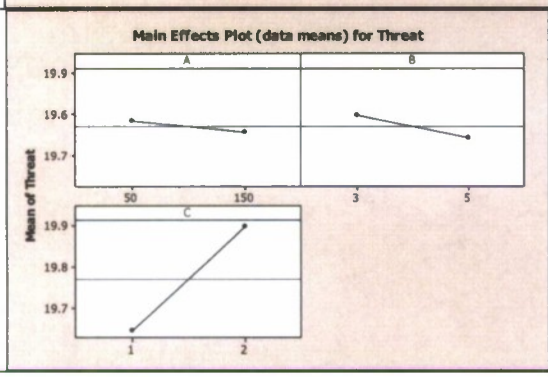




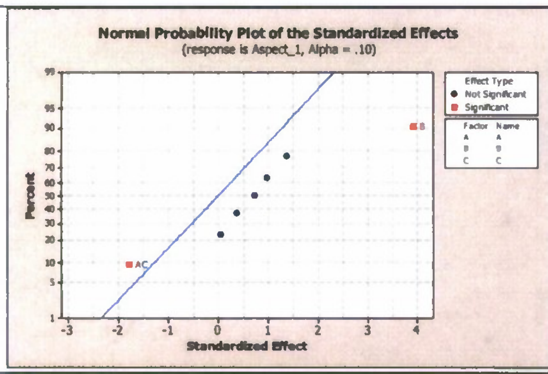
Track to Track Threat Consistency:

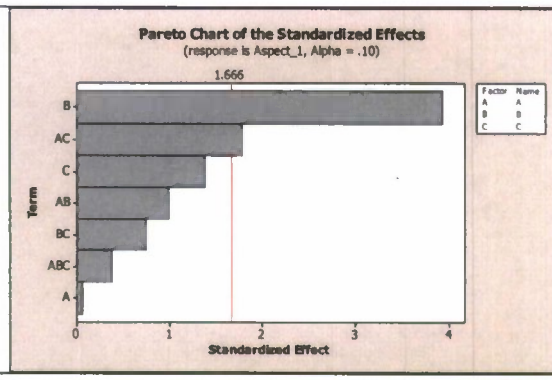


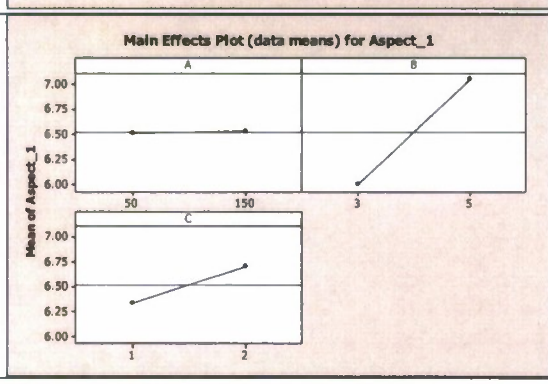




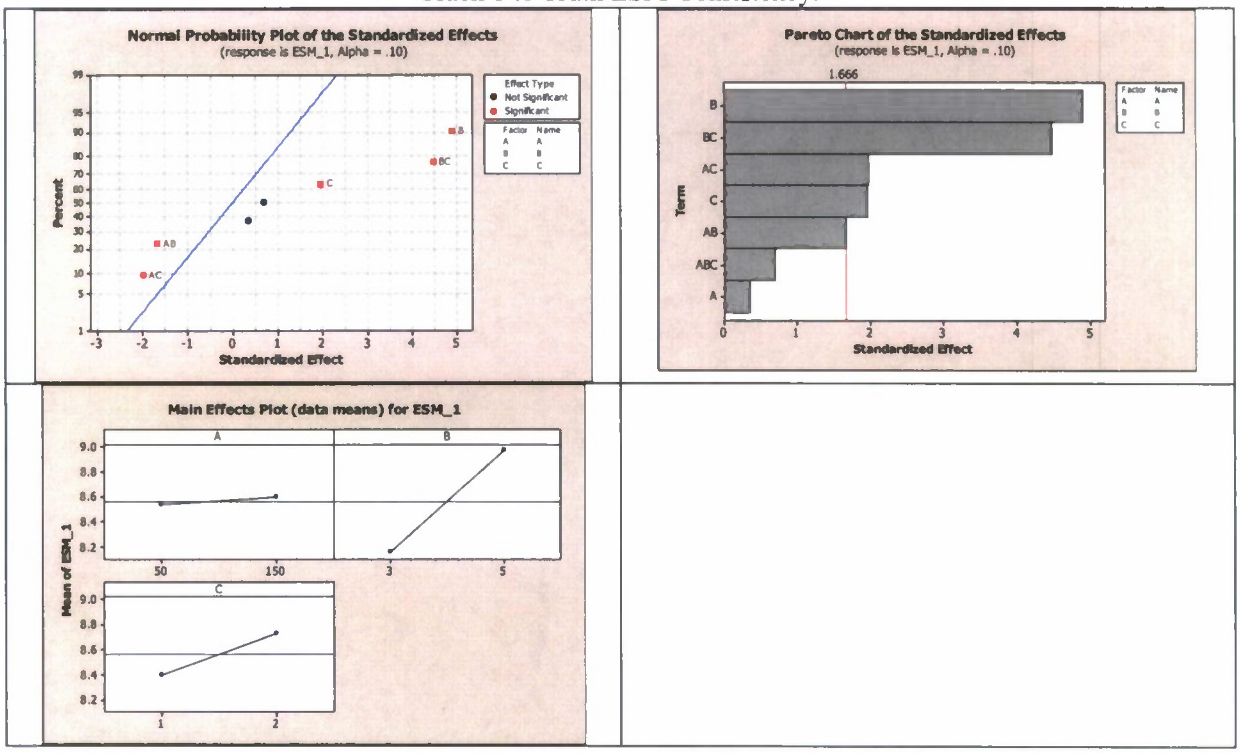
Track 1 to Truth Aspect Consistency:



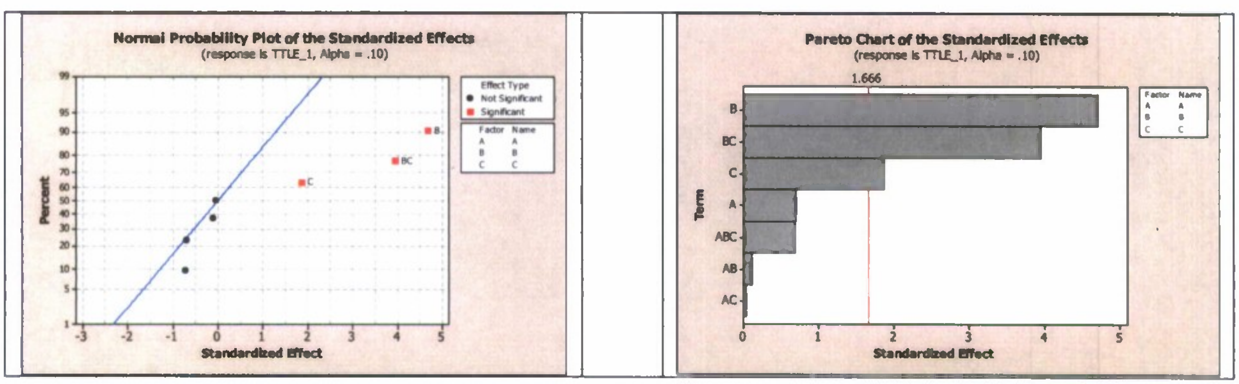


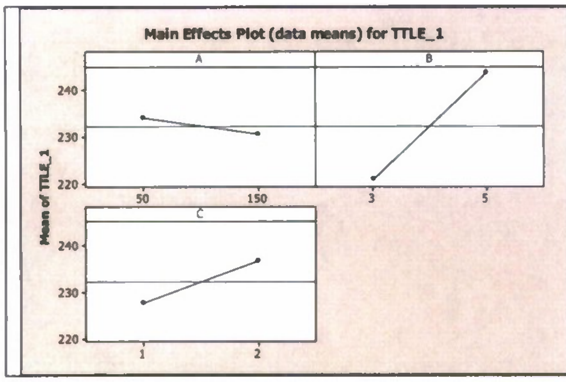


Track 1 to Truth ESM Consistency:

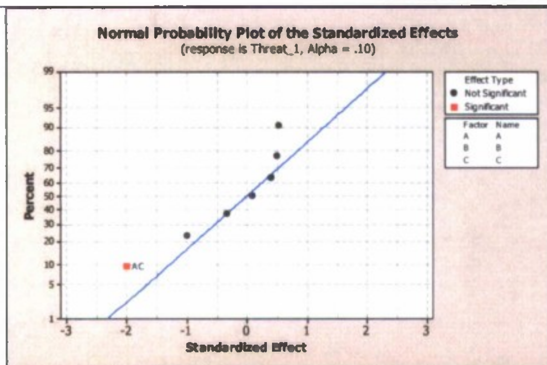


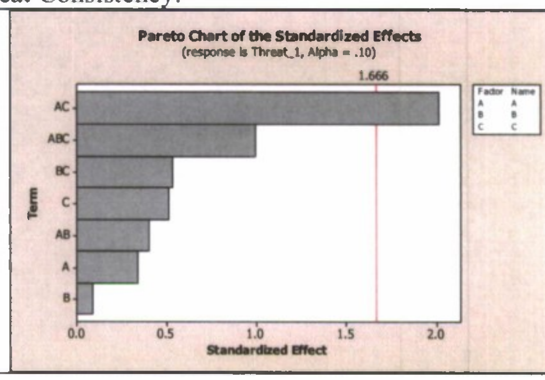
Track 1 to Truth TTLE Consistency:

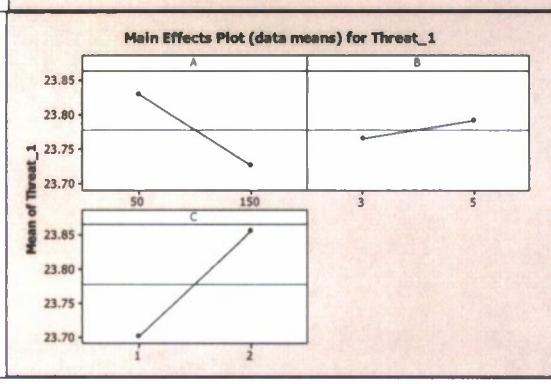




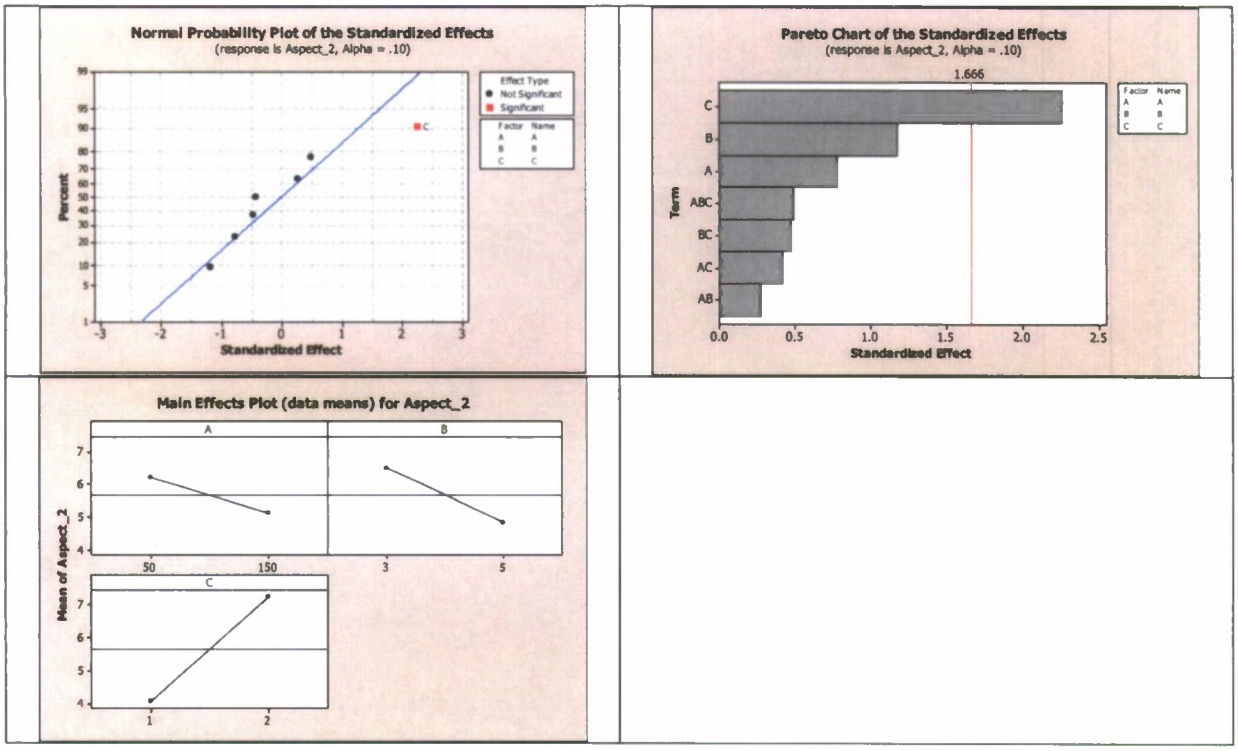
Track 1 to Truth Threat Consistency:



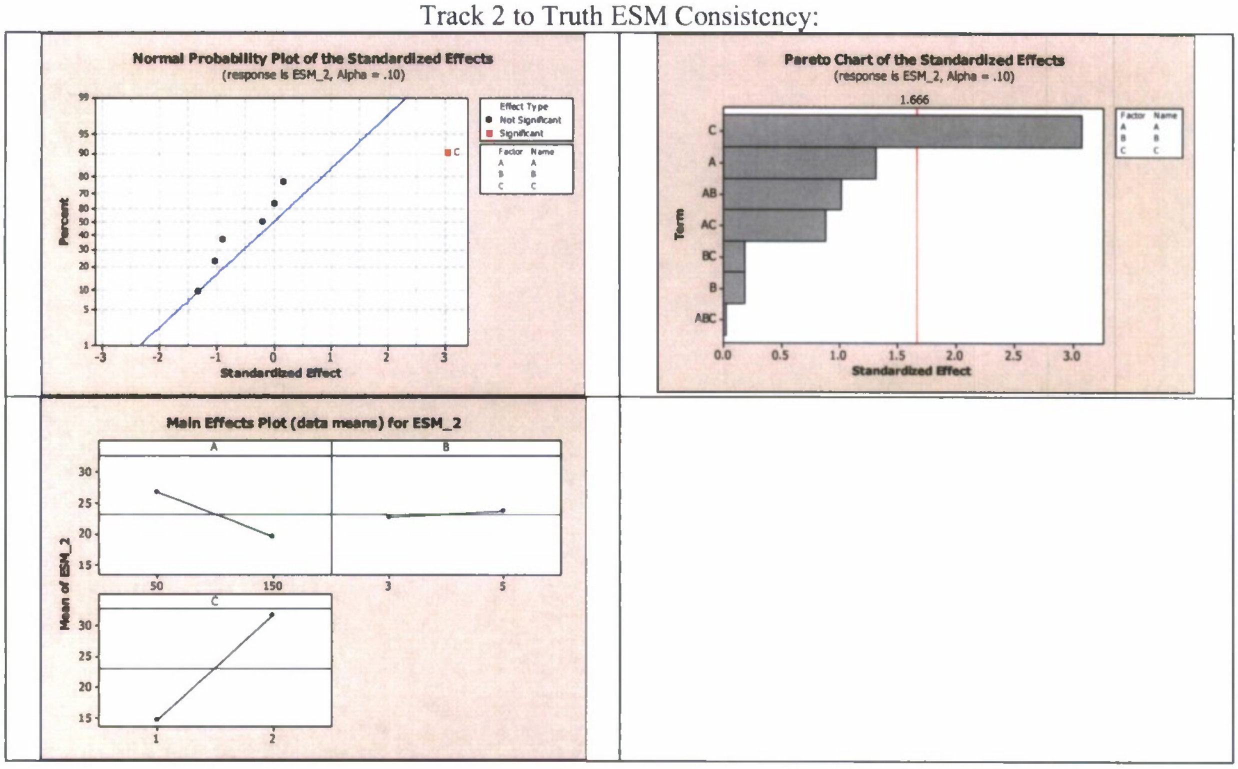




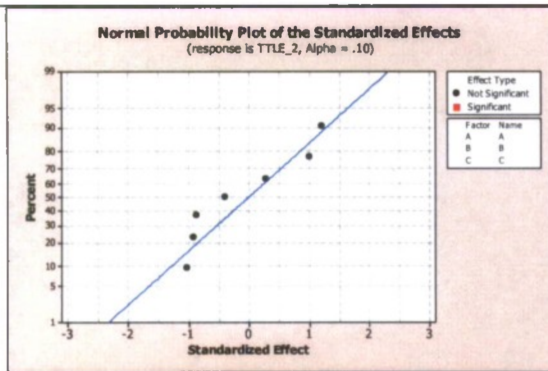
Track 2 to Truth Aspect Consistency:

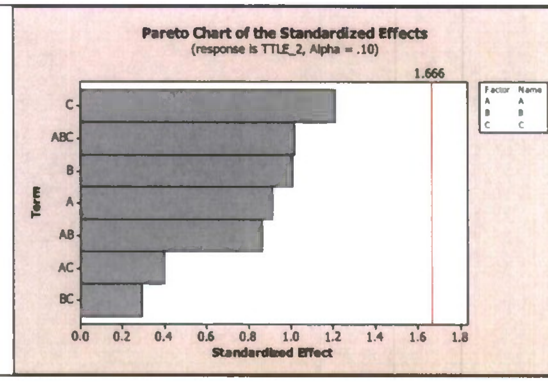


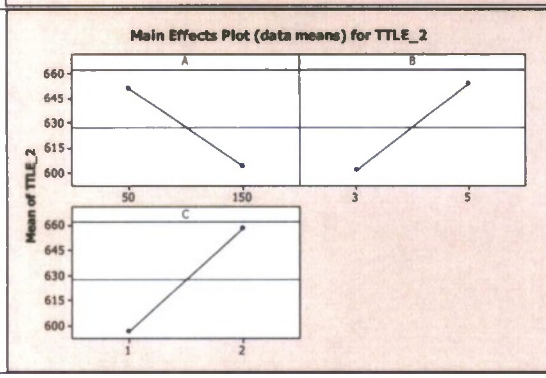
Track 2 to Truth ESM Consistency:



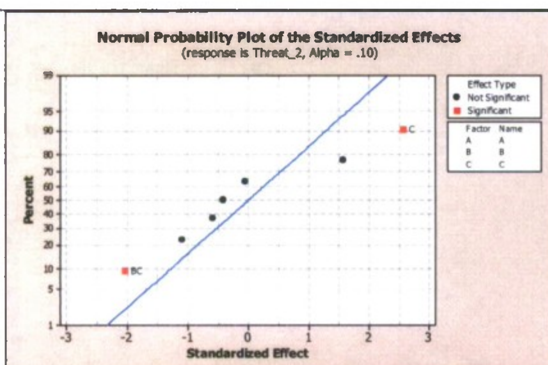
Track 2 to Truth TTLE Consistency:

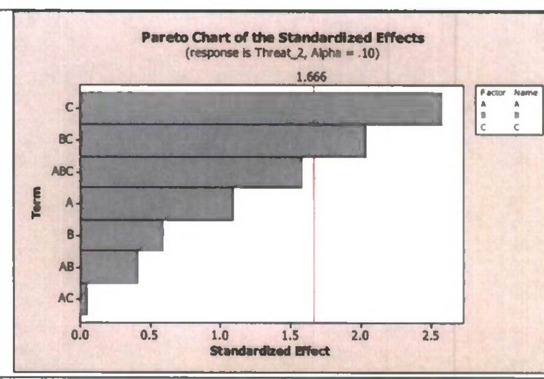


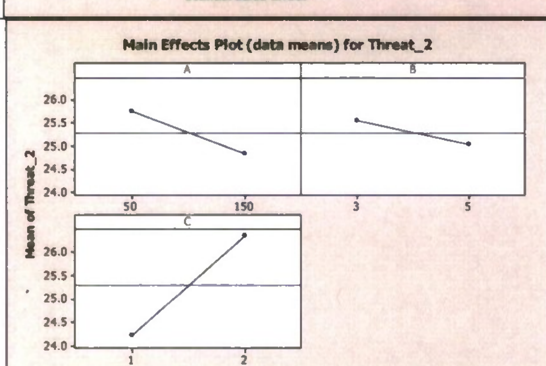




Track 2 to Truth Threat Consistency:





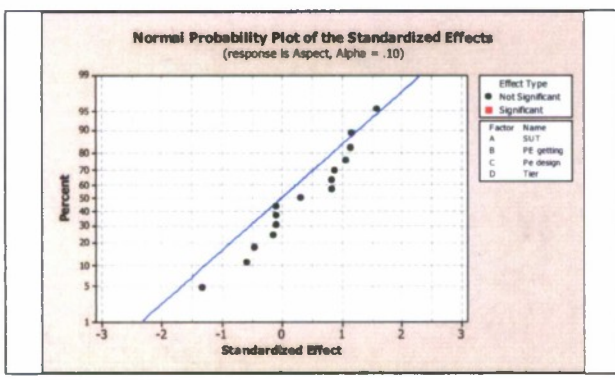


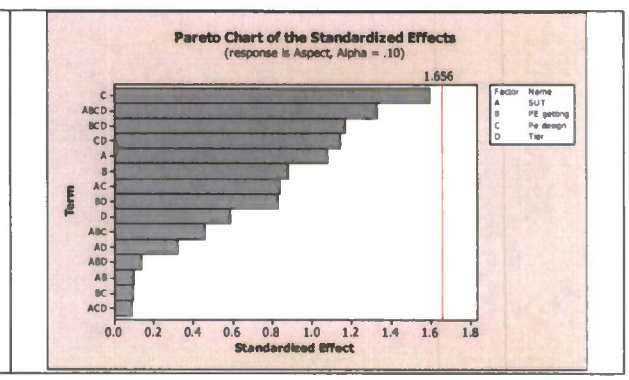
D. L2 and L3 Inter Tier DOE Charts

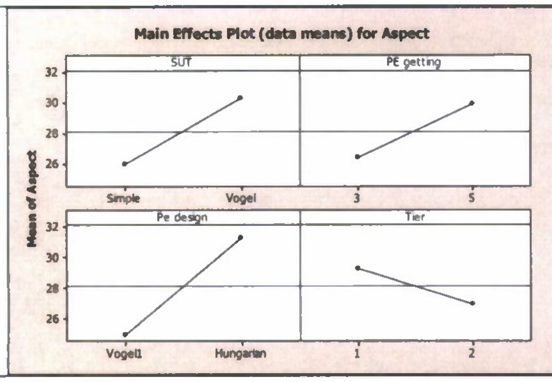
This section provides the Inter Tier DOE charts conducted in Section 5. The four factors SUT Design Gating Factor, PE Gating Factor, PE Design and Tier Level at two levels each are tested to find which of these factors affect the MOPs significantly. In Inter Tier we have three sensors on 2 platforms and they fuse data within and across platforms (Tier 2) and just within platform (Tier 1). So we have to analyze track-to-truth and track-to-track associations for each of the MOPs. The summary of the results is shown in Table 6. Here for each MOP we have the Normal Probability plot which summarizes the significant factors. Then for the significant factors we plot the main effects plot which tells us how the change in factor affects the MOP. For the significant interactions we plot the interaction plot which shows the effect of change in factor level combination on MOP.

After taking a look at the summary Table 6, we can say that factor "Inter Tier" is comparatively more significant than PE Gating Factor, PE Design and SUT Design Gating Factor. In this case the DOE table suggests that fusing data across platforms reduces the discrepancies in the input data.

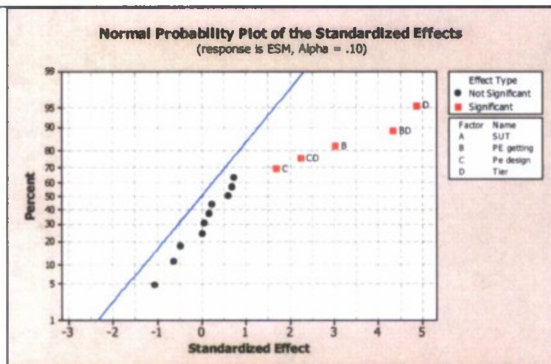
Track To Track Aspect Consistency:

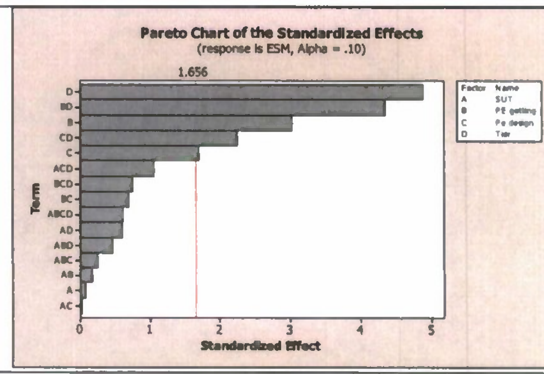


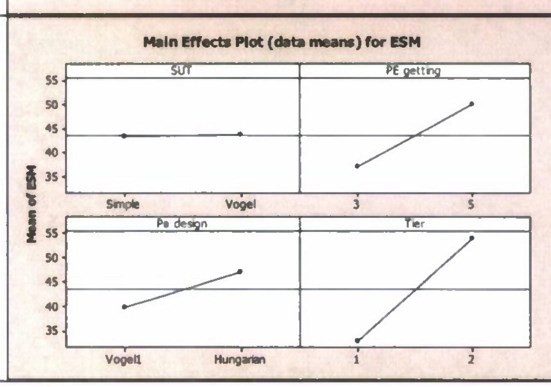




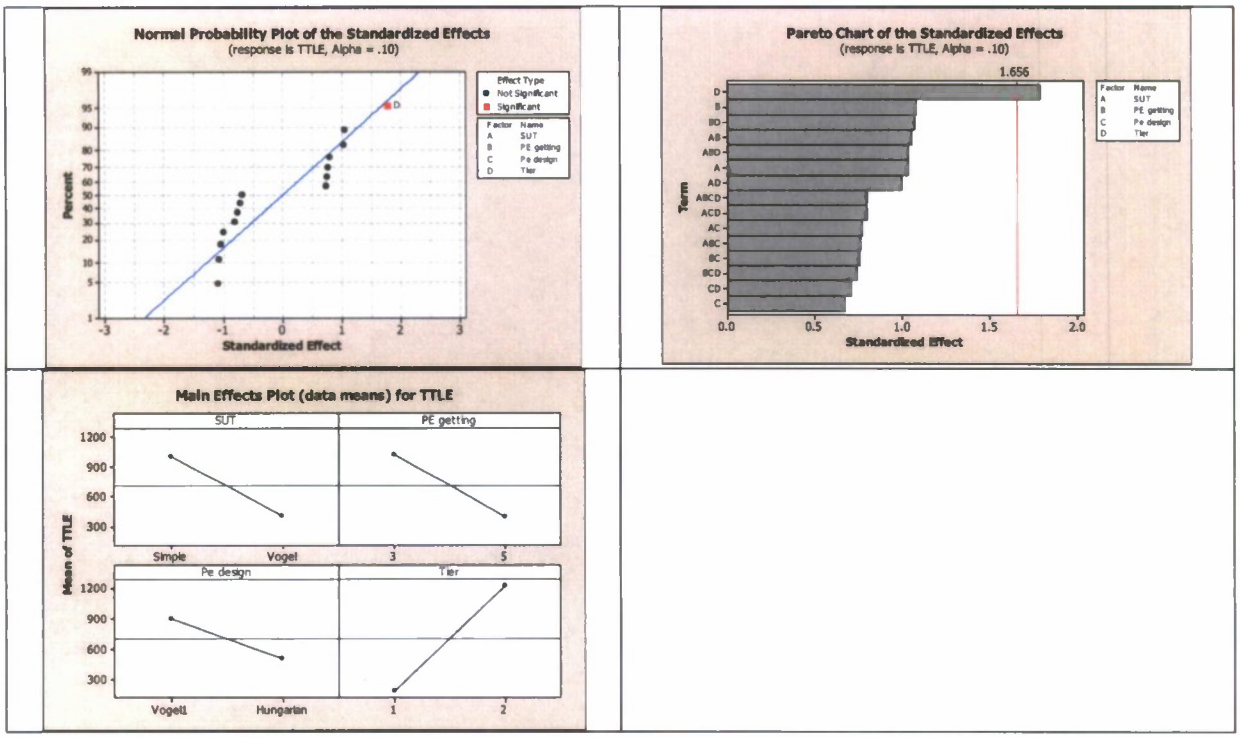
Track To Track ESM Consistency:



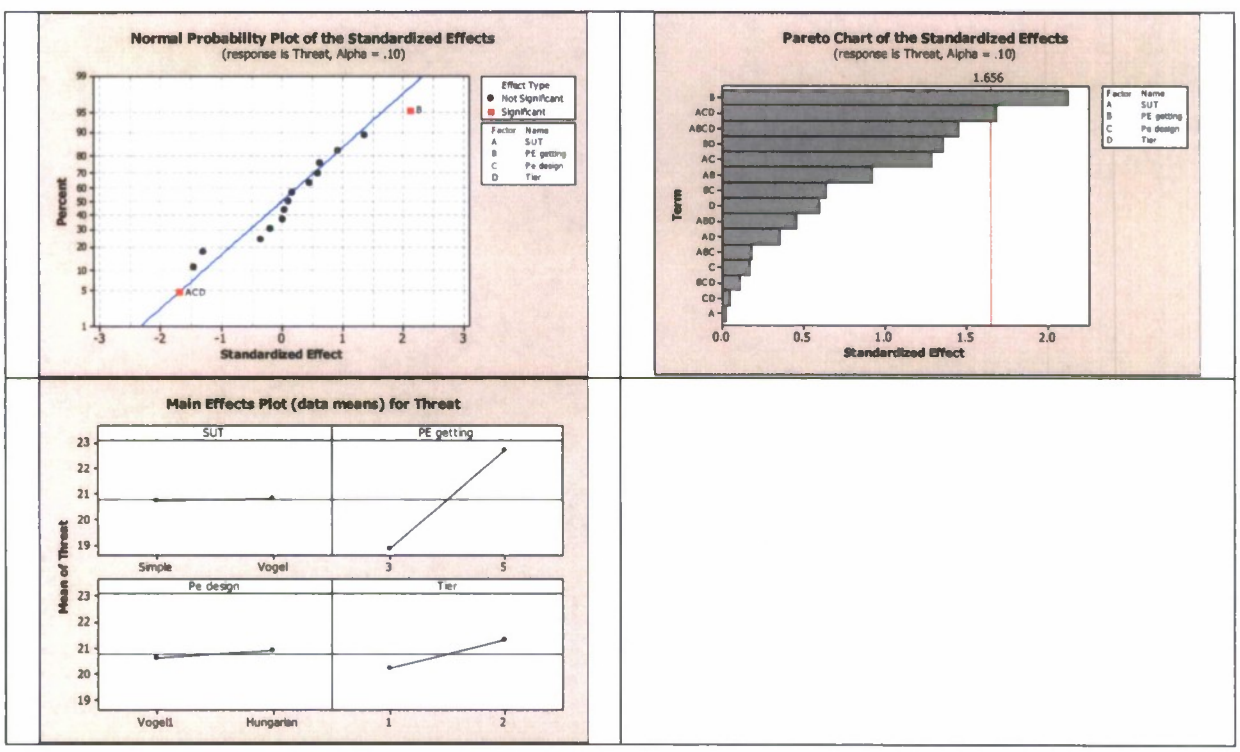




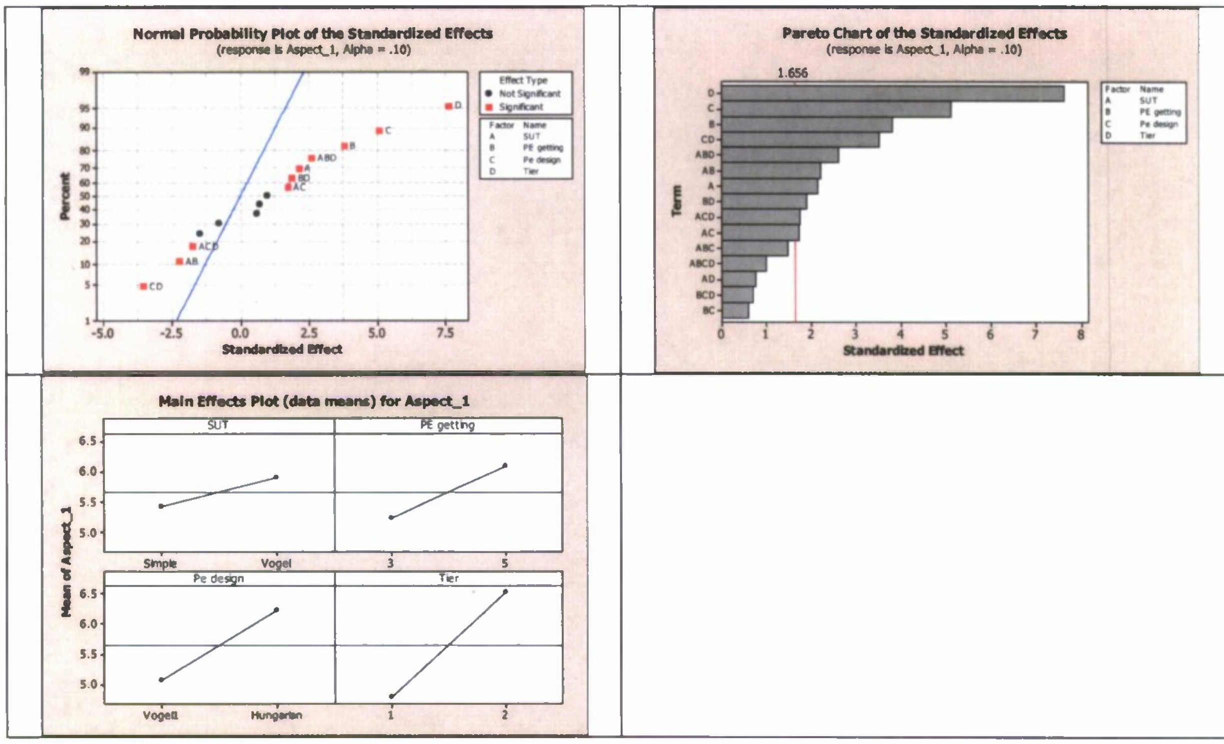
Track To Track TTLE consistency:



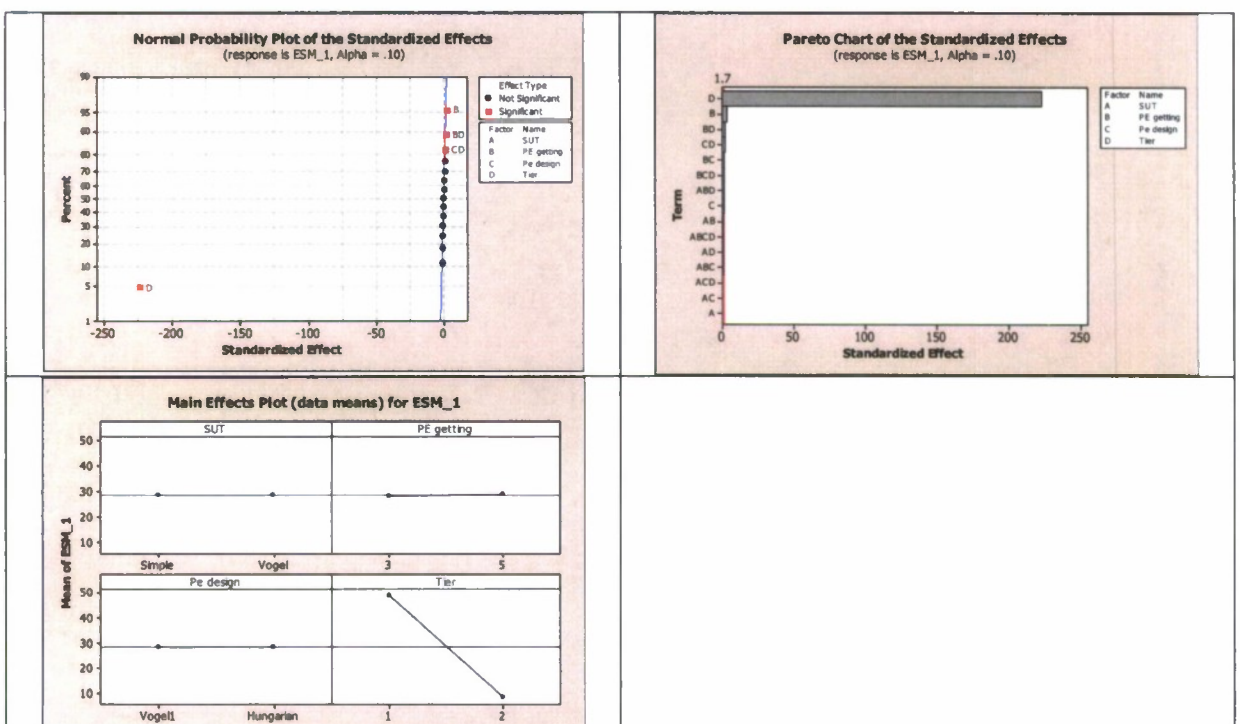
Track To Track Threat Consistency:



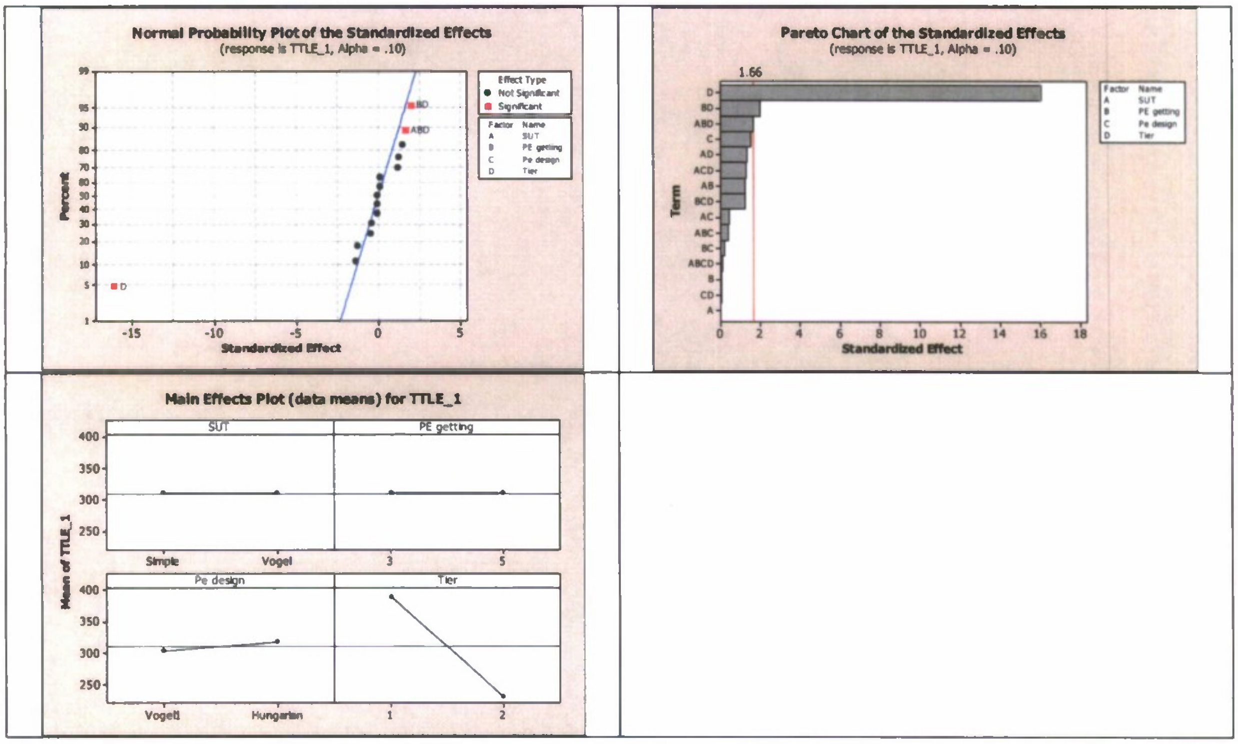
Track 1 to Truth Aspect Consistency:



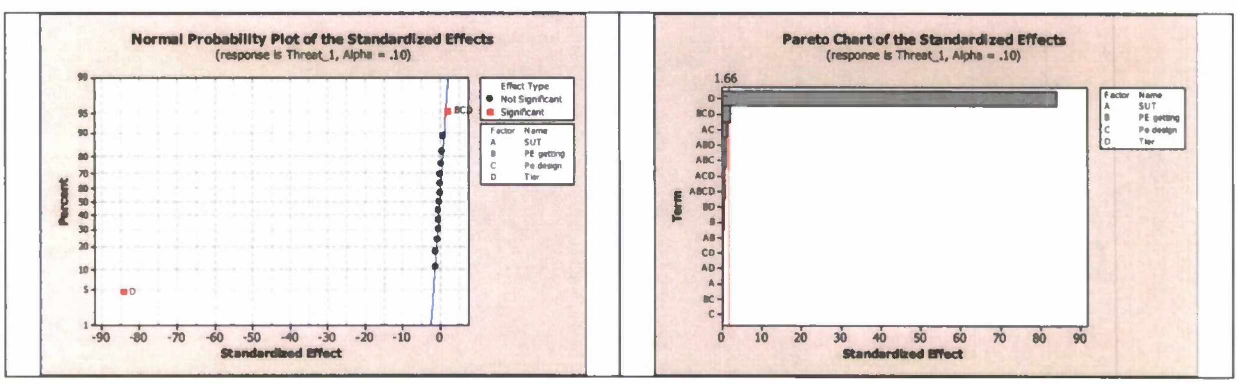
Track 1 to Truth ESM Consistency:

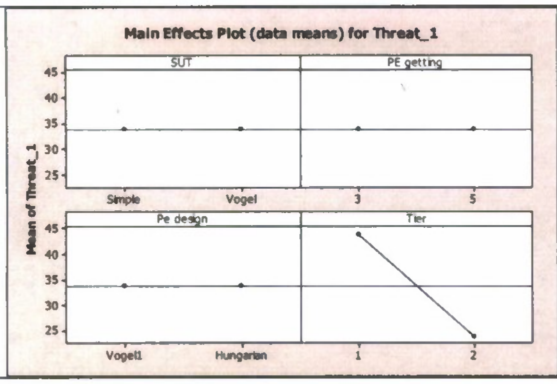


Track 1 to Truth TTLE Consistency:

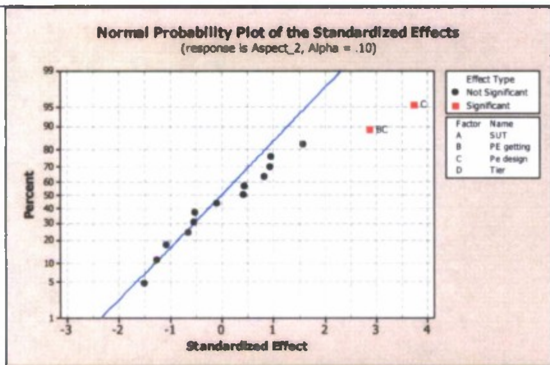


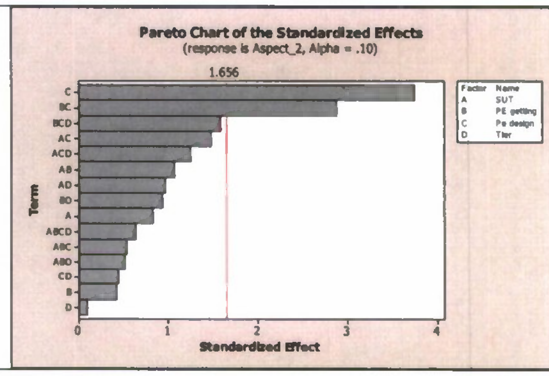
Track 1 to Truth Threat Consistency:

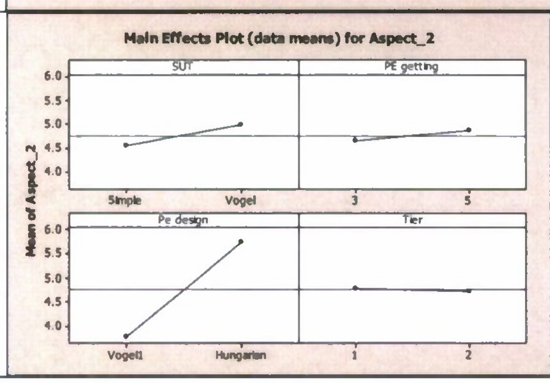




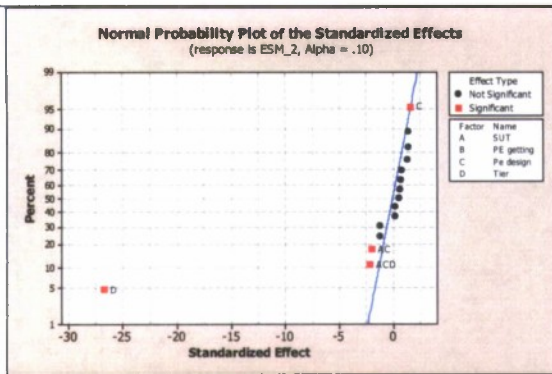
Track 2 to Truth Aspect Consistency:

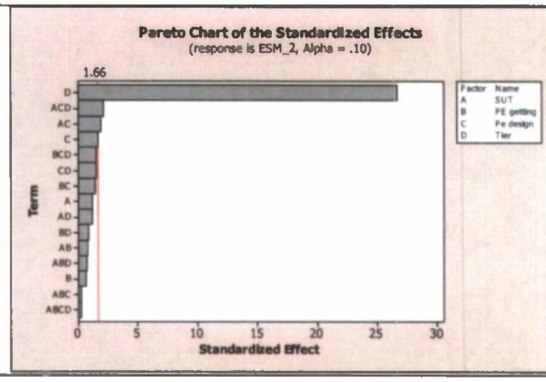


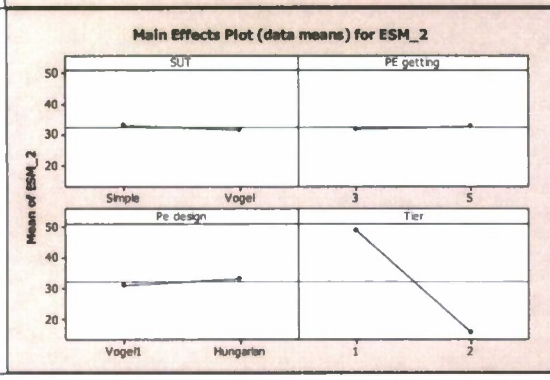




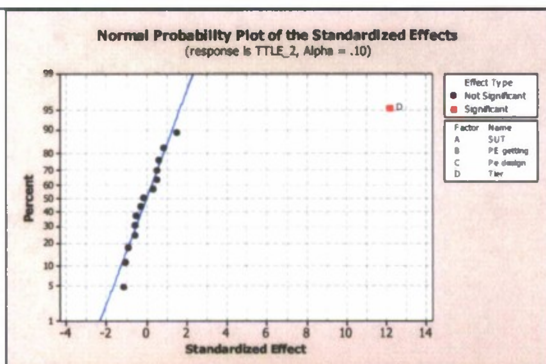
Track 2 to Truth ESM Consistency:

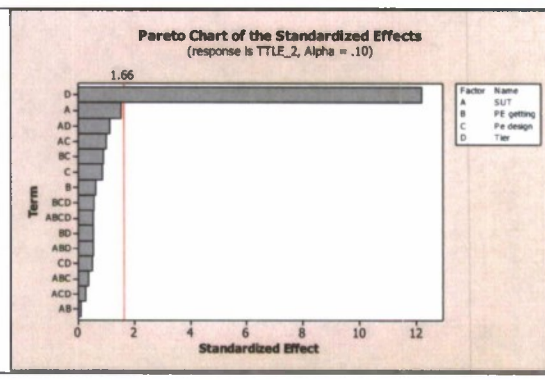


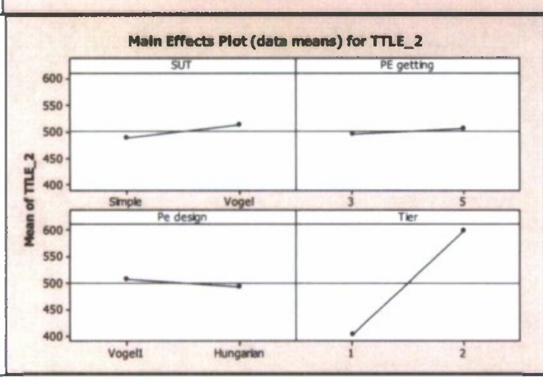




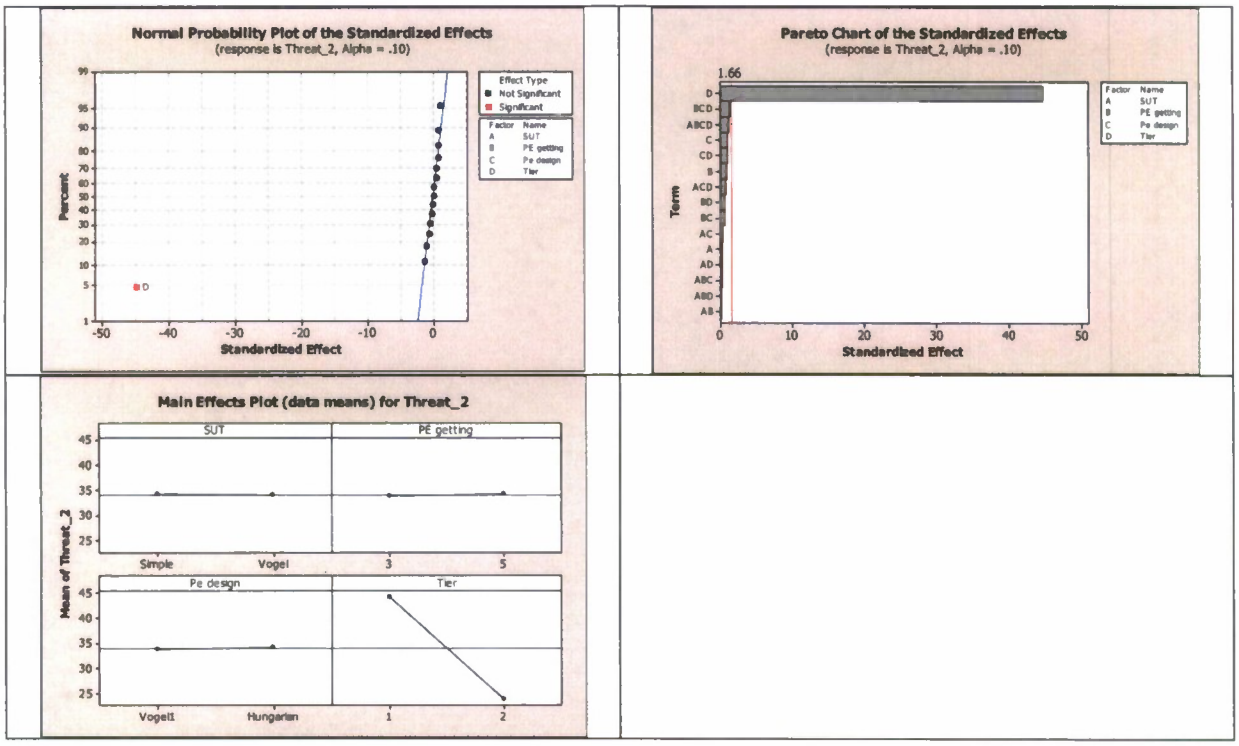
Track 2 to Truth TTLE Consistency:







Track 2 to Truth Threat Consistency:

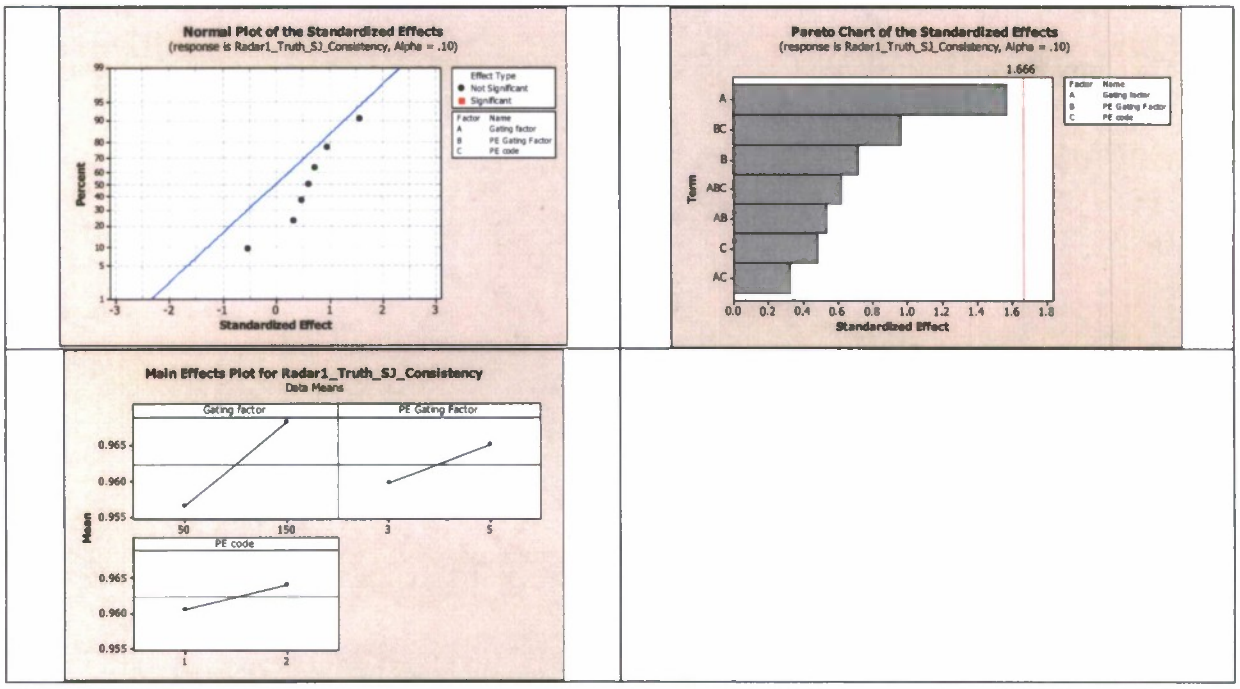


E. L4 Tier 0, 1, 2 and Inter Tier DOE Charts

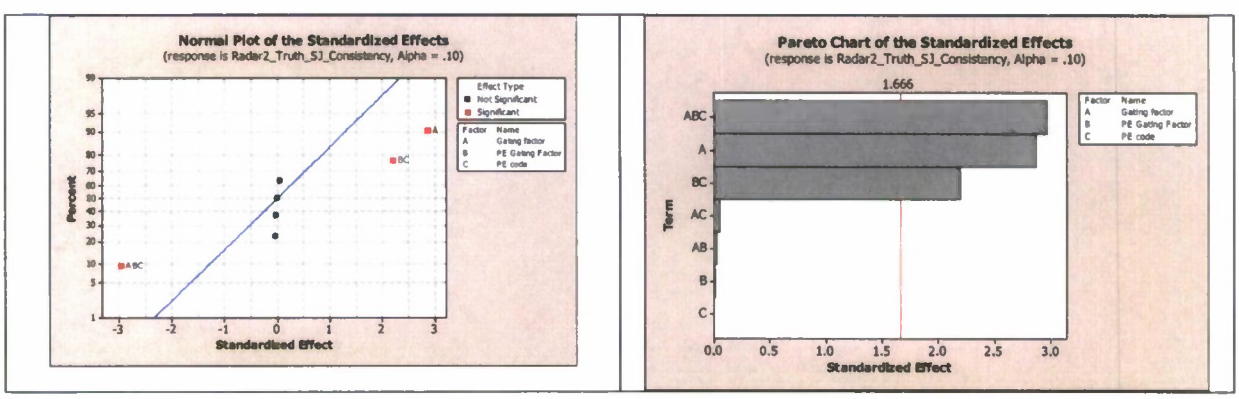
This section provides the Inter Tier DOE charts conducted in Section 5. The four factors SUT Design Gating Factor, PE Gating Factor, PE Design and Tier Level at two levels each are tested to find which of these factors affect the MOPs significantly. In Inter Tier we have three sensors on 2 platforms and they fuse data within and across platforms (Tier 2) and just within platform (Tier 1). So we have to analyze track-to-truth and track-to-track associations for each of the MOPs. The summary of the results is shown in Table 6. Here for each MOP we have the Normal Probability plot which summarizes the significant factors. Then for the significant factors we plot the main effects plot which tells us how the change in factor affects the MOP. For the significant interactions we plot the interaction plot which shows the effect of change in factor level combination on MOP.

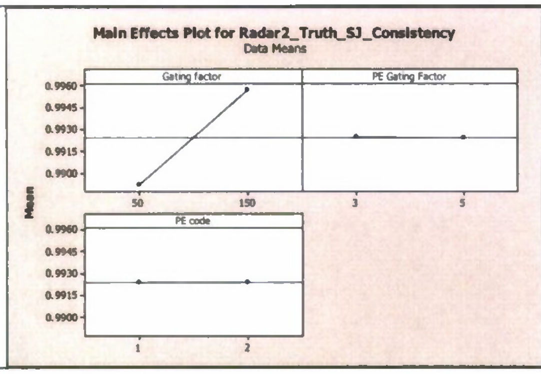
After taking a look at the summary Table 6, we can say that factor "Inter Tier" is comparatively more significant than PE Gating Factor, PE Design and SUT Design Gating Factor. In this case the DOE table suggests that fusing data across platforms reduces the discrepancies in the input data.

Tier 0 Radar 1 to Truth S/J ration Consistency

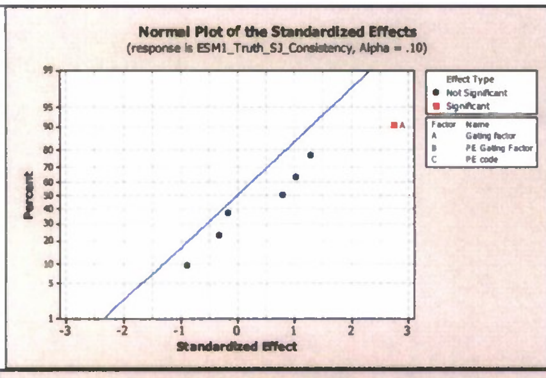


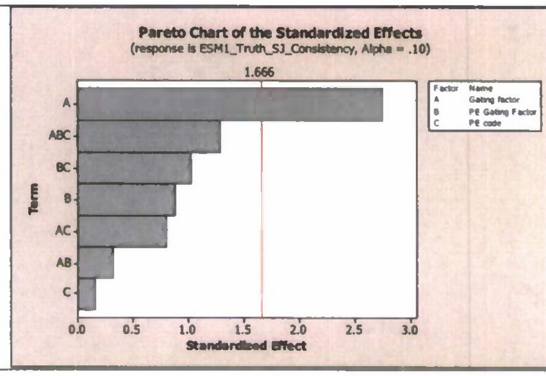
Tier 0 Radar 2 to Truth S/J ration Consistency:

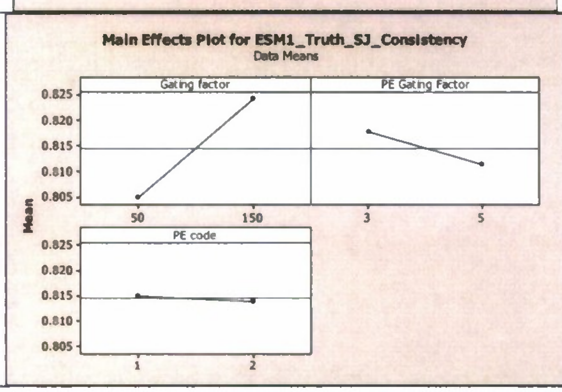




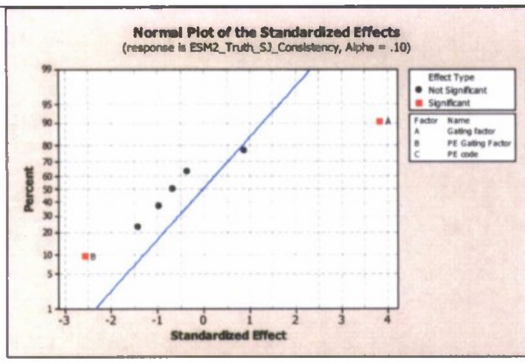
Tier 0 ESM 1 to Truth S/J Ratio Consistency:

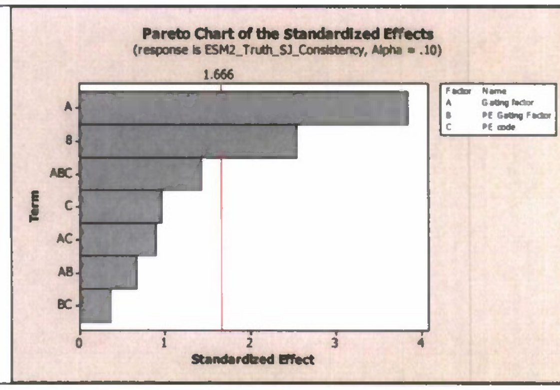


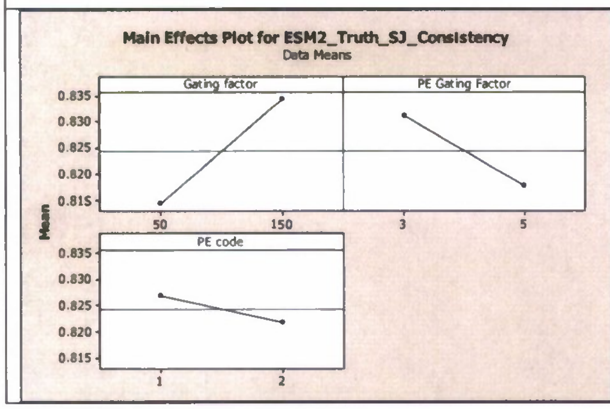




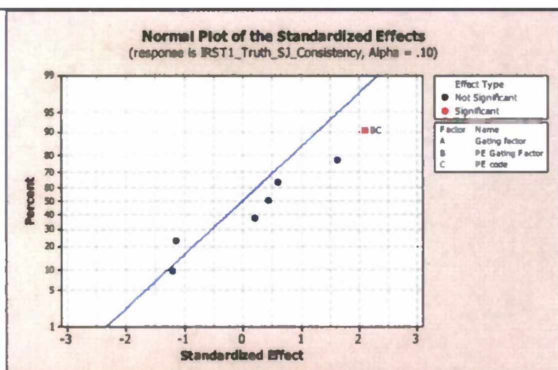
Tier 0 ESM 2 to Truth S/J Ratio Consistency:

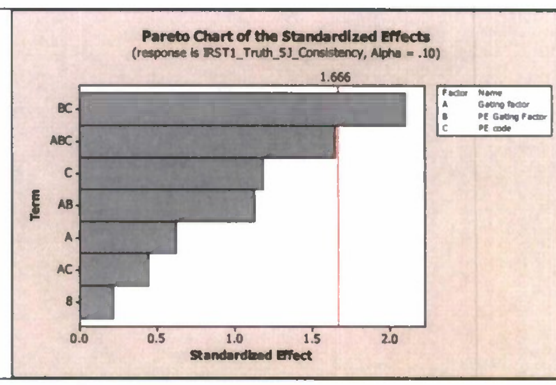


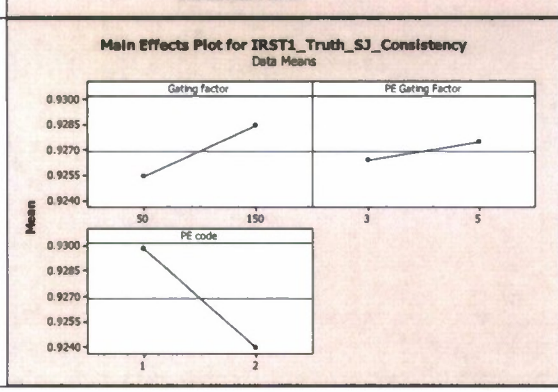




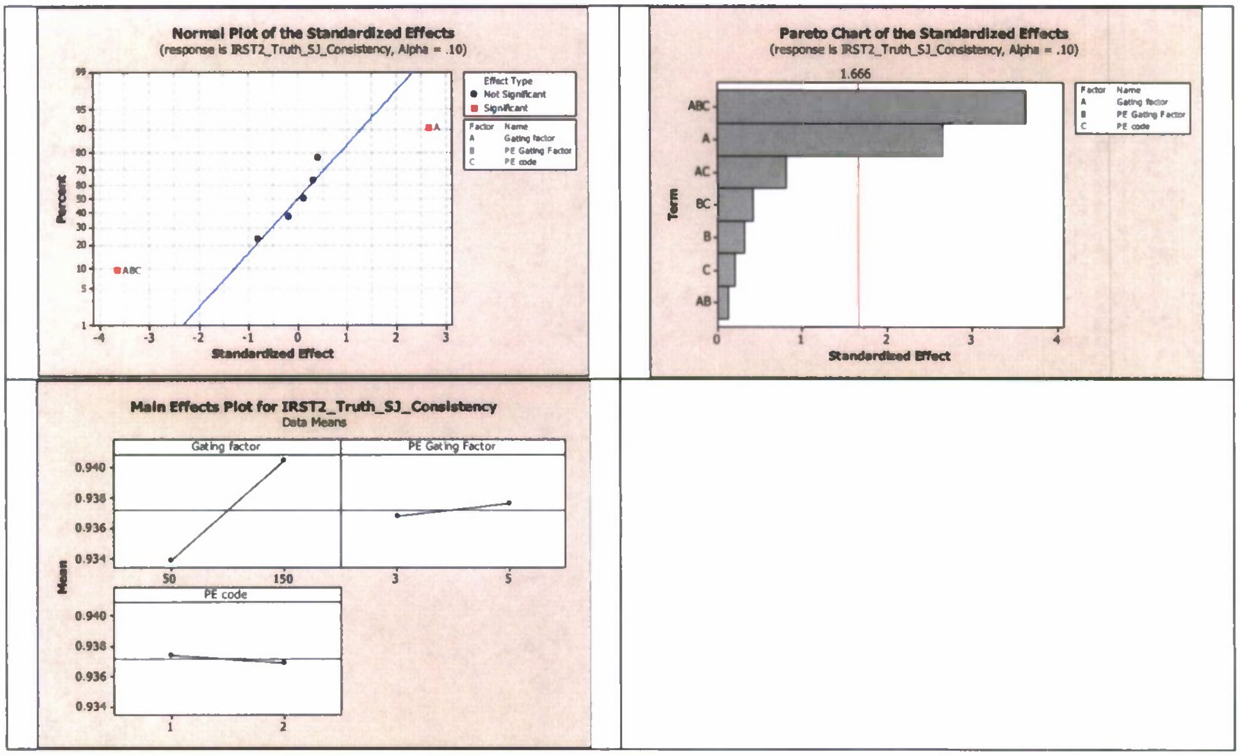
Tier 0 IRST 1 to Truth S/J ration Consisteney:



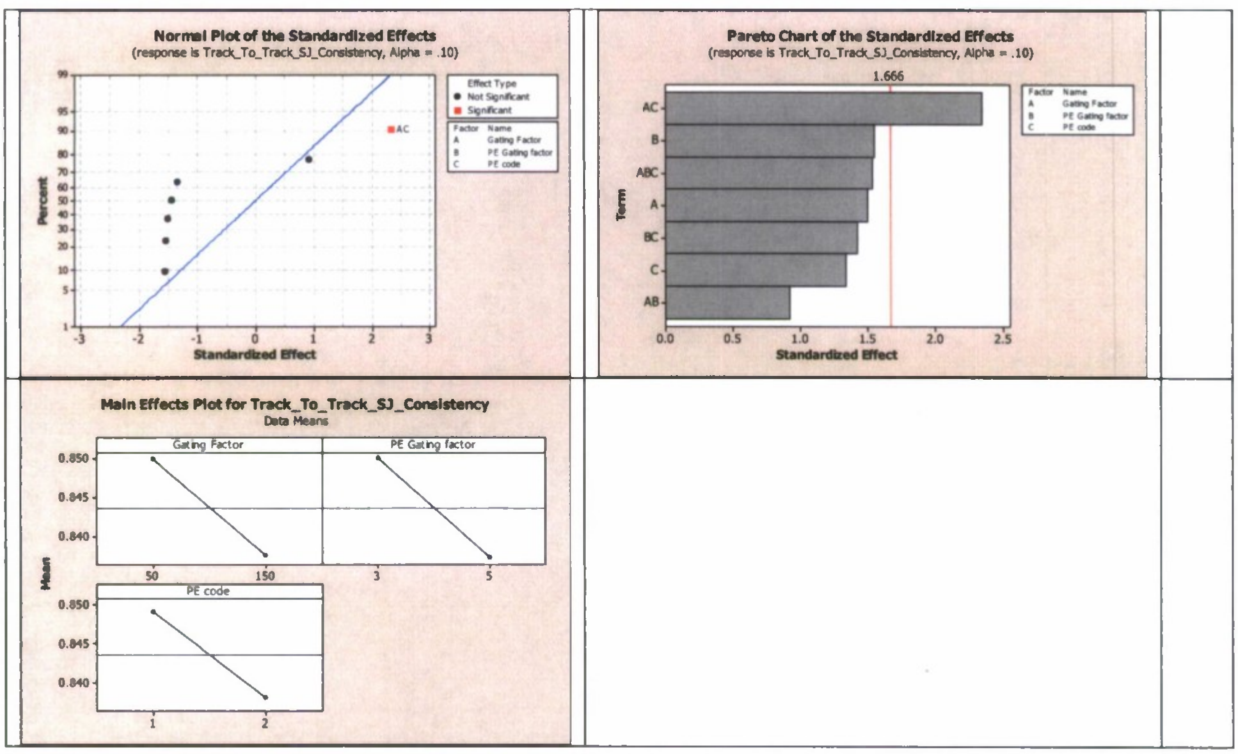




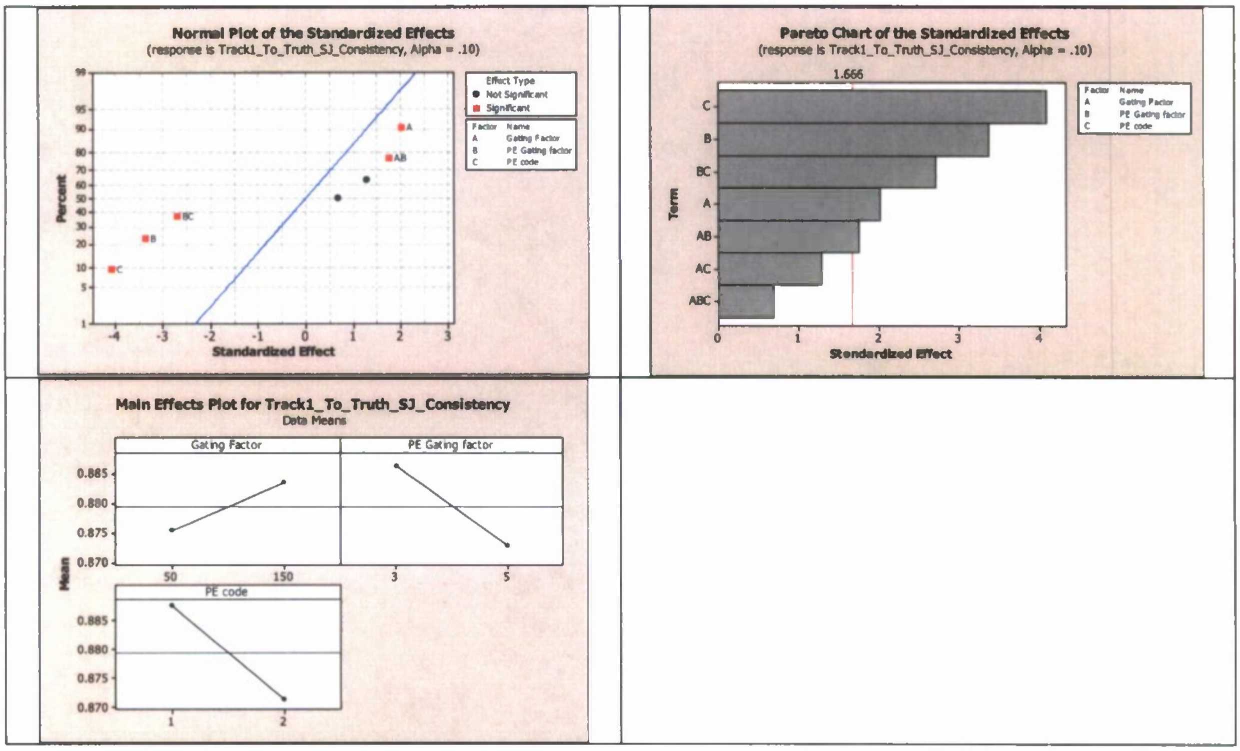
Tier 0 IRST 2 to Truth S/J ration Consistency:



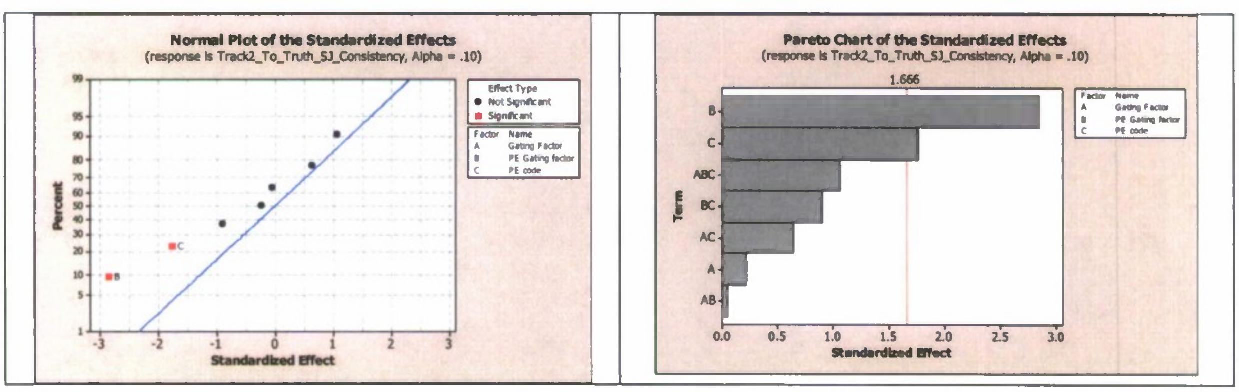
Tier 1 Track To Track S /J Consistency:

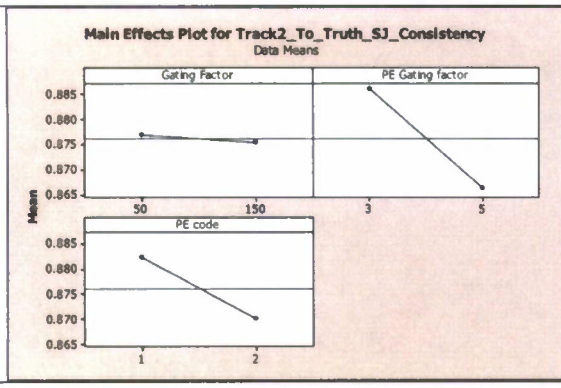


Tier 1 Track 1 To Truth S/J Consistency:

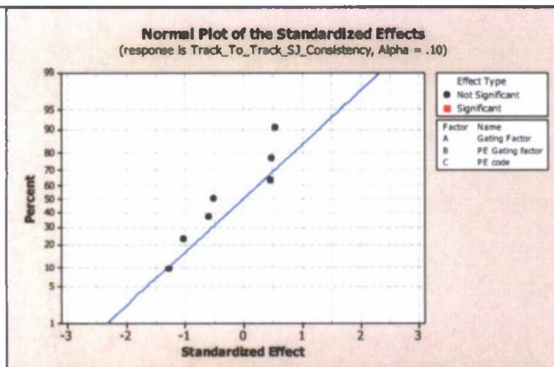


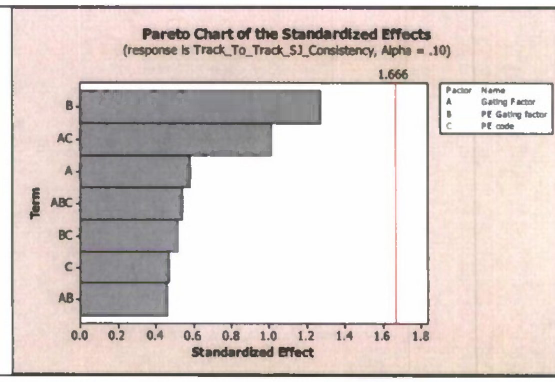
Tier 1 Track 2 To Truth S/J eonsistency:

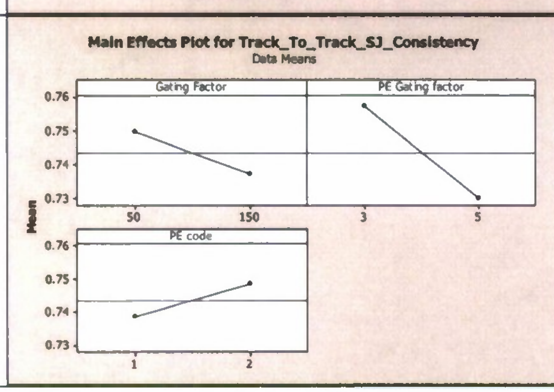




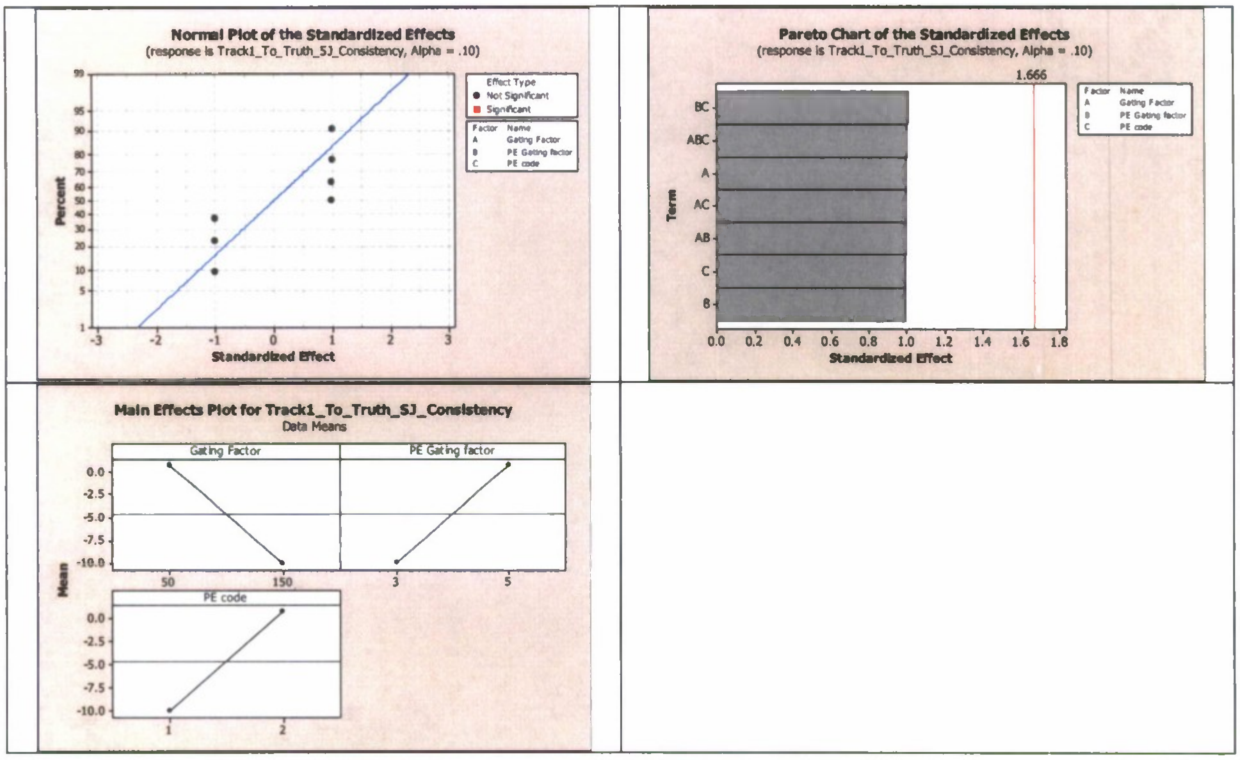
Tier 2 Track to Track S/J Consistency:



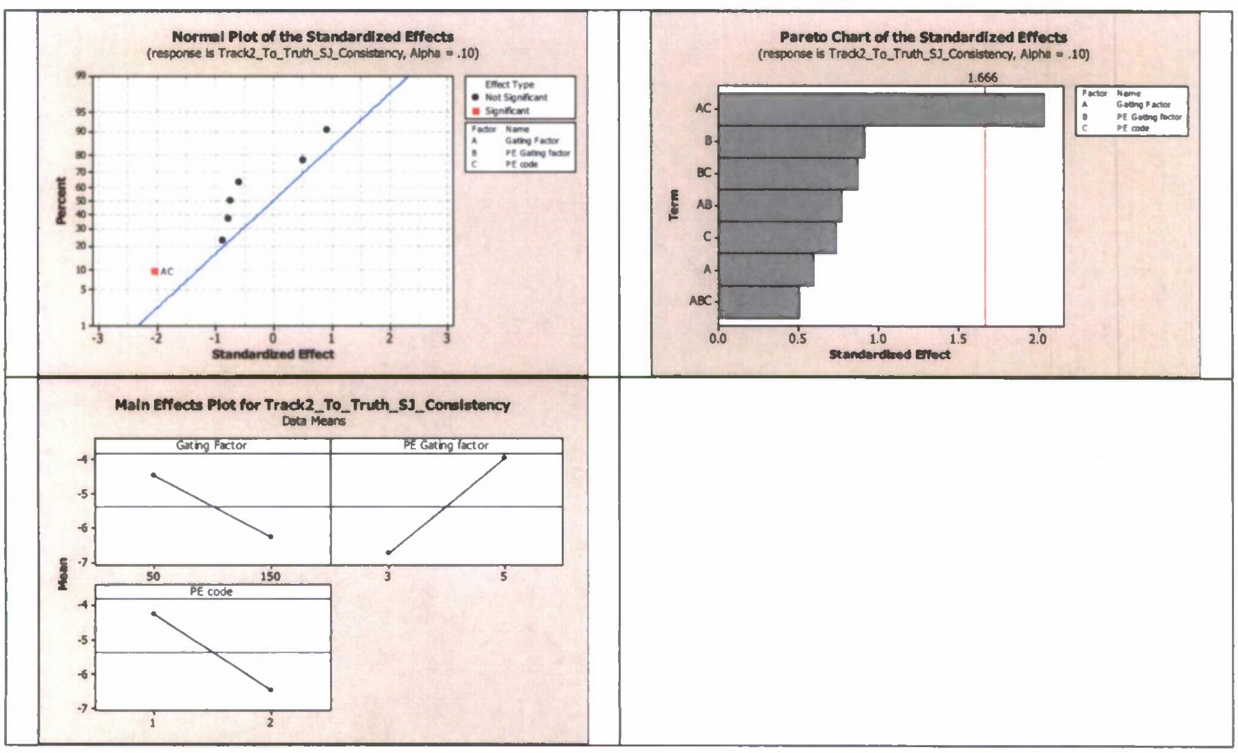




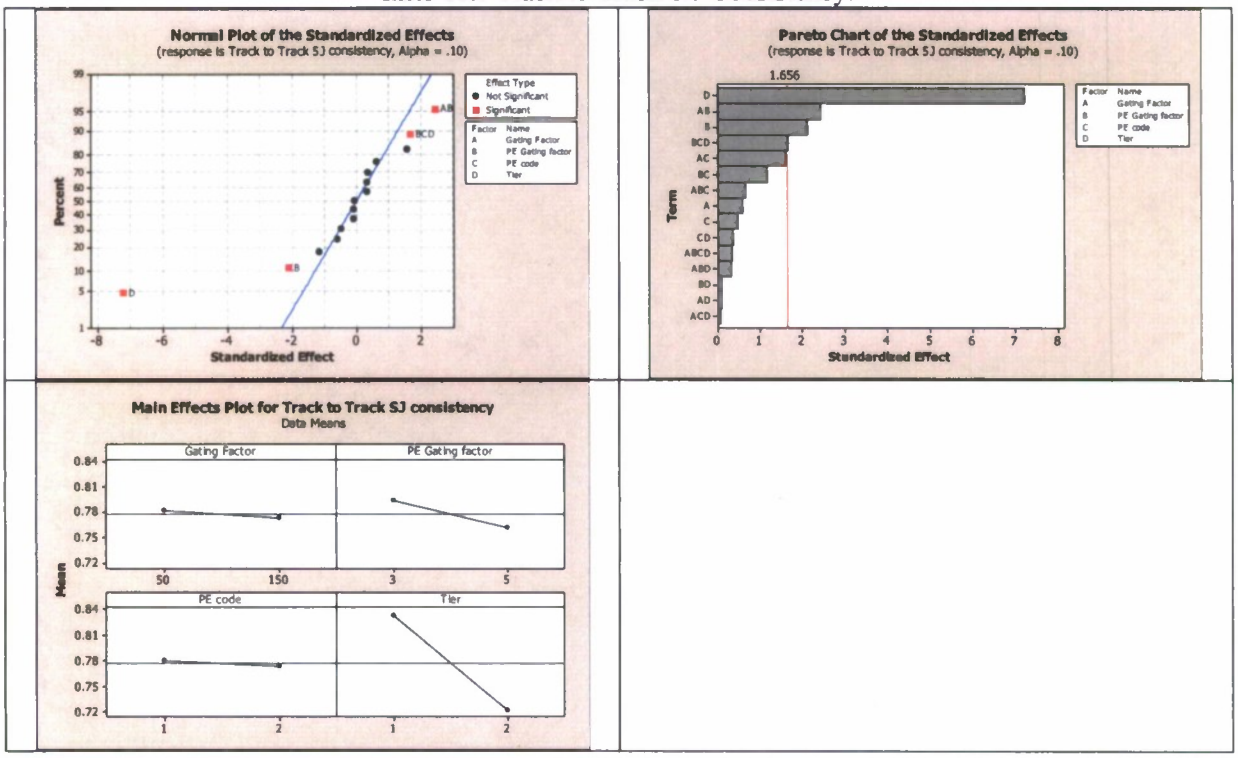
Tier 2 Track 1 To Truth S/J Consistency:



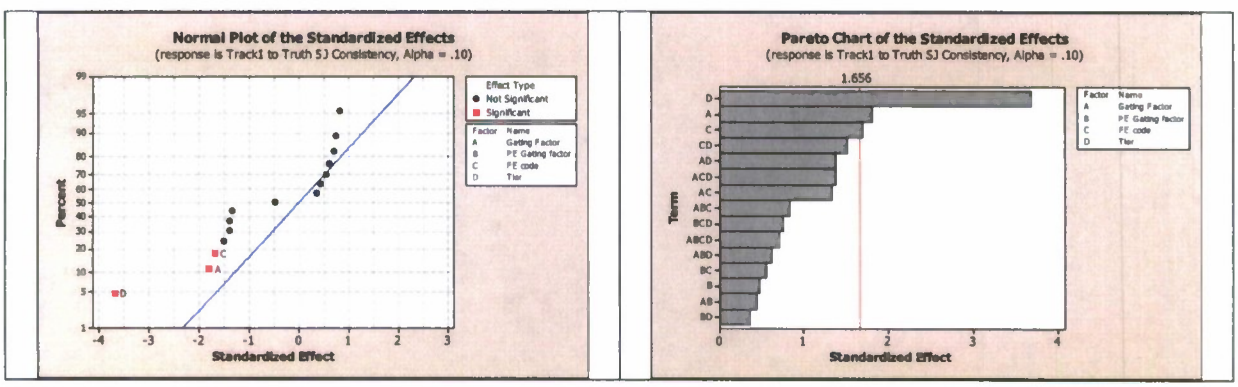
Tier 2 Track 2 to Truth S/J Consistency:

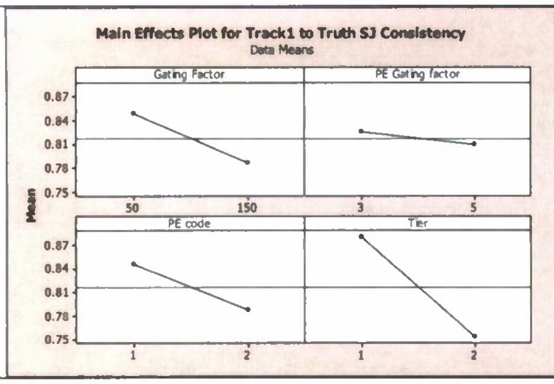


Inter Tier Track to Track S/J Consistency:



Inter Tier Track 1 to Truth S/J Consistency:





Inter Tier Track 2 to Truth S/J Consistency: