

A BAYESIAN FRAMEWORK TO INTEGRATE KNOWLEDGE-BASED AND DATA-DRIVEN INFERENCE TOOLS FOR RELIABLE YIELD DIAGNOSES

Chih-Min Fan

Yun-Pei Lu

No. 135 Yuan-Tung Road
Department of Industrial Engineering and Management, Yuan Ze University
Taoyuan, TAIWAN, R.O.C.

ABSTRACT

This paper studies the issues of designing a Bayesian framework for the reliable diagnosis of various yield-loss factors induced in semiconductor manufacturing. The proposed framework integrates both the results from knowledge-based and data-driven inference tools as input data, where the former resembles expert's knowledge on diagnosing pre-known yield-loss factors while the latter serves for exploring new yield-loss factors. Three modules with specific designs for yield diagnosis applications are addressed: Pre-Process for generating candidate factors and corresponding prior distributions, Bayesian Inference for calculating posterior distributions, and Post-Process for deriving reliable rankings of candidate factors. The final output, a Bubble Diagram with Pareto Frontier, provides visual interpretations on the integral results from data-driven, knowledge-based and Bayesian inference tools. Specific issues addressed in the proposed Bayesian framework provide directions for implementing a real system.

1 INTRODUCTION

In semiconductor manufacturing industry, with the shorter product lifecycle, the capability to fulfill customer's Time-To-Market (TTM) request for profit maximization determines the competitive advantage of a fab. One of the keys to guarantee TTM is the rapid yield ramp-up, which relies on the quick diagnosis of yield-loss factors induced by hundred steps of manufacturing processes. Nevertheless, as technology involves into the sub-wavelength era, not only the number of yield-related parameters rapidly increases but also the relationship among these parameters become more complex than before. Rapid yield diagnosis for realizing TTM becomes a tough challenge to the semiconductor manufacturing fab.

A typical yield diagnosis procedure is demonstrated in Figure 1. As can be seen, the diagnosis procedure is triggered by a yield fault event and followed by two abstract phases: find fault symptom and diagnose root cause. The

migration from the first phase to the second phase is through a critical function named as first-cut diagnosis. Due to the fact that the output of first-cut diagnosis is not 100% reliable, the suspected cause may not be the root cause of yield losses. As a result, there are always a lot of iterations for root-cause validations. An effective yield diagnosis procedure could then be defined as the procedure with (1) the minimum iterations of loop L1 for the quick root-cause validation and (2) the minimum iterations of loop L2 for the thorough root-cause validation. It is concluded that the function of first-cut diagnosis is critical to the performance of yield diagnosis procedures.

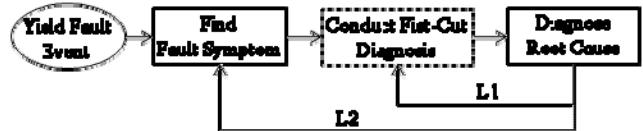


Figure 1: A typical yield diagnosis procedure

To improve the function of first-cut diagnosis, there are two approaches and corresponding systems developed in industry practices: knowledge-based and data-driven inferences. Most of the knowledge-based inference applications are rule-based expert systems. The rule-based expert system generates a priority list of suspected root causes with fault possibilities, which resemble the decision results made by a group of yield analysis engineers. For example, the combined use of fuzzy rule modeling, fuzzy inference, and neural network learning for first-cut diagnosis has been published in (Guo et al. 1996). On the other hand, with the huge volume of engineering data and advances of modern computing power, the data-driven inference applications are also widely applied to first-cut diagnosis. For example, the data-mining technique adopts a sequence of ANOVA test or decision tree analysis (Chien, Wang, and Cheng 2007) to generate a priority list of suspected root causes based on their statistical significances. Both data-driven or knowledge-based inference tools have their own pros and cons.

The knowledge-based inference tool can produce a satisfactory priority list of suspected root causes when the real root cause is similar to other cases happened before. However, due to the fact that all inference rules are either input by engineers or learned from historical yield-loss cases, it is difficult to infer the new type of yield-loss cases beyond the scope of engineer's current knowledge. In a modern fab, it is not allowed that a yield loss case similar to the old cases happens again. As a result, in current practices, the knowledge-based expert systems are just used for training purpose instead of serving as an assistant tool to diagnose the new problems.

On the contrary, the data-driven inference could search a wide range of suspected root causes beyond the engineer's current knowledge. Moreover, each suspected root cause is based on its statistical significance, which is an objective evidence demonstrated by the data. Nevertheless, the statistical empirical model behind the data-driven inference to generate the statistical significance may not really match to the underlying yield model. And, very often, the data-driven method itself is not robust to various data characteristics and situations. As a result, usually, the root cause is not ranked as the top orders of the priority list. Instead, the factor ranked as the top orders are usually false-identified and even domain-nonsense at all.

Neither knowledge-based nor data-driven inference tools alone performs well in the first-cut diagnosis function. Under this situation, engineers will take a lot of time and efforts on the loops L1 & L2 defined in Figure 1. To enhance the performance of the first-cut diagnosis function, it raises a problem: how to develop a quick solution to combine both the advantages of knowledge-based and data-driven inference tools? A totally new methodology and system built from the scratch is one of the strategies. But what is it? How much efforts and how long will it take to develop such a new methodology and system? How to make sure of its success? What can we do before such an idea methodology and system is ready there?

This paper tries to improve the performance of first-cut diagnosis by integrating and reusing the existed systems and knowledge in industry practices. Instead of building a totally new methodology and system from the scratch, this research will integrate and reuse the existed knowledge-based and data-driven inference tools. In specific, a Bayesian framework is proposed and expected to generate a reliable ranking list for the effective first-cut diagnosis. The overall architecture is demonstrated in Figure 2. It consists of three modules: Data-Driven Inference, Knowledge-Based Inference and Bayesian Ranking Inference. The former two modules are existed systems in industry practices, while the third module is totally new and is the focus of this paper.

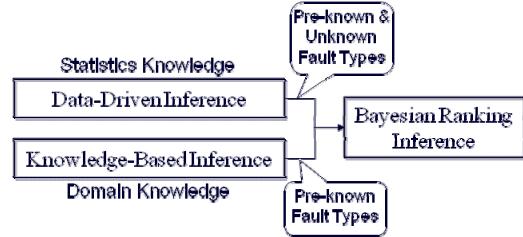


Figure 2: System reuse and integration for effective first-cut diagnosis

2 PROBLEMS & SOLUTION FRAMEWORK

2.1 Problems of Data-Driven Inference Techniques

Almost all the data-driven inference techniques adopted by semiconductor manufacturing suffers two problems: False Identification Due to Confounding Variables and Miss-Identification Due to One-Factor-At-A-Time Analysis. Detailed explanations and corresponding examples are described follows.

False Identification(FI) Due to Confounding Variables

To illustrate the problem of directly applying data-driven inference tools such as ANOVA or Regression Test to test the differences among equipments, Figure 3 illustrates a simulation example. In this example, there are k steps ($\text{Step}_1, \text{Step}_2, \dots, \text{Step}_k$), in which the equipment $\text{EQP}_{1,B}$ in Step_1 is simulated as abnormal. It is expected that the minimum p-value will occur at Step_1 after applying ANOVA or Regression Test. However, the test result is that Step_k has the smallest p-value. Thus Step_1 is miss-identified while Step_k is false-identified.

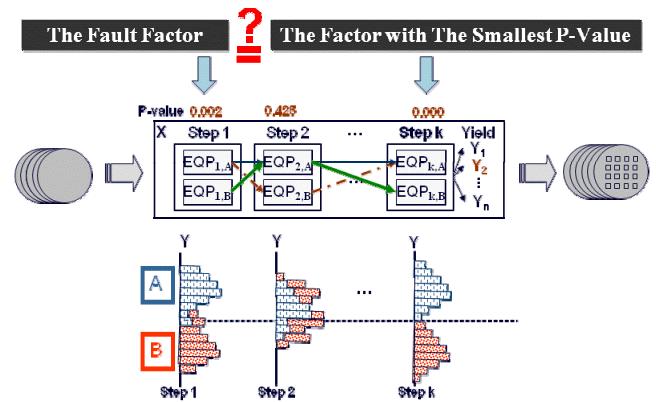


Figure 3: False identification due to confounding variables

The false identification of $\text{EQP}_{k,B}$ in Step_k as the abnormal equipment is due to the misunderstanding that the

step with the smallest p-value is the faulty step. The right interpretation should be:

- The $\text{EQP}_{k,B}$ in Step_k could be the abnormal equipment, or
- The $\text{EQP}_{k,B}$ in Step_k could be normal but is falsely identified as abnormal due to the confounding variable $\text{EQP}_{1,B}$ in Step₁, i.e. most abnormal wafers with low yield processed by the faulty equipment $\text{EQP}_{1,B}$ in step 1 also go through the normal equipment $\text{EQP}_{k,B}$ in Step_k.

Whether the wafers are processed by $\text{EQP}_{k,A}$ or $\text{EQP}_{k,B}$ at step k is determined by various scheduling rules. As the number of process steps increases, it is more likely that there will have process steps falsely identified as abnormal.

Miss-Identification(MI) Due to One-Factor-At-A-Time Analysis

To demonstrate the miss-identification due to one-factor-at-a-time analysis, another simulation is conducted as the second example. Figure 4 illustrates the corresponding data, which is simulated by the regression model

$$y = \mu_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_a. \quad (1)$$

where μ_0 is the mean of wafer yield, β_j is the yield loss due to the abnormal equipment $\text{EQP}_{j,B}$ in Step_j, $X_j=1$ if the wafer is processed by $\text{EQP}_{j,B}$, $X_j=0$ if the wafer is processed by $\text{EQP}_{j,A}$, and ε randomly varies from wafer to wafer.

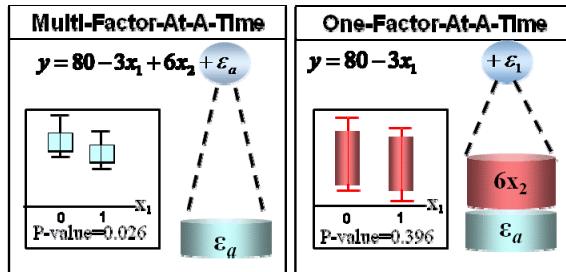


Figure 4: Miss-identification due to one-factor-at-a-time analysis

If we only apply the one-factor-at-a-time analysis to X_2 , under the assumption that X_i is independent with each other, the full regression model in (1) will be reduced to

$$y = \mu_0 + \beta_1 x_1 + \varepsilon. \quad (2)$$

Compare (1) with (2), it can be concluded that $\varepsilon = \beta_2 x_2 + \varepsilon_a$, i.e. the regression-error ε of β_1 under one-factor-at-a-time analysis involves the interference from other factors and is much larger than the regression-

error ε_a estimated by multi-factor-at-a-time analysis.

Thus, due to the larger regression-error ε , the β_1 estimated by one-factor-at-a-time analysis will be less significant and could be miss-identified. The box plot with red dashed-line in Figure 4 provides the visual interpretation of miss-identification due to one-factor-at-a-time analysis.

2.2 Bayesian Framework

Due to the problems FI & MI, the data-driven inference results becomes unreliable. As the number of explanatory factors increases, the impacts of FI & MI are much more significant. In a modern semiconductor manufacturing fab, there could be more than one thousand factors to be investigated. The capability of current data-driven inference techniques for yield diagnosis is seriously doubted.

The development of solutions to cope with the problems FI & MI has two perspectives: Algorithm and System. From the algorithm perspective, all data-driven data mining algorithms suffer the same problem as FI. One of the possible ways to solve FI is to apply domain knowledge so that the false-identified yield-loss factors could be filtered out. On the other hand, Multi-Factor-AT-A-Time analysis is an intuitive approach to solve MI. Thus an advanced Bayesian model selection algorithm with Multi-Factor-At-A-Time analysis is further developed.

From the system perspective, recognizing the importance of reusing existed tools and systems in industry practices, we proposed to integrate both knowledge-based and data-mining inference tools for solving FI & MI (Figure 2), where

- the knowledge-based inference results resemble the provider of domain knowledge for Bayesian inference, and
- the data-driven inference tool generates the set of candidate factors for Multi-Factor-AT-A-Time analysis.

Figure 5 demonstrates the system framework for Bayesian Model Selection. There are three modules: Pre-Processing, Bayesian Inference and Post Processing. Issues of each module are described in the following sections.

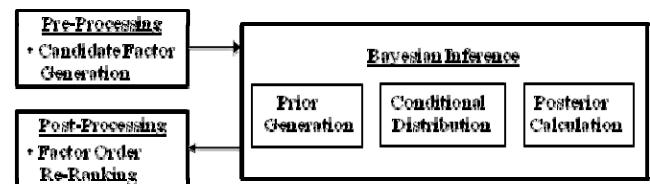


Figure 5: Framework of Bayesian Ranking Inference

3 PRE-PROCESSING FOR BAYESIAN INFERENCE

There are a lot of factors generated from the data-driven and knowledge-based inference tools respectively. The factors from data-driven tools could include the new factors unidentified before, but they also suffer from the FI and MI problems. On the contrary, the factors from knowledge-based systems could be free from FI and MI, but they may not include the new factors not identified before.

Too many factors will diminish both the quality and efficiency of Bayesian inference. There is a need to conduct factor filtering for a smaller set of candidate factors. There are two stage of filtering. As shown in Figure 6, with the setting of thresholds for statistical p-value and subjective fault possibility, the factors unlikely to be faulty are filtered out first. Then the second stage of factor filtering is achieved by selecting the final candidate factors from the intersection of the candidate factors from both data-driven tools and knowledge-based systems.

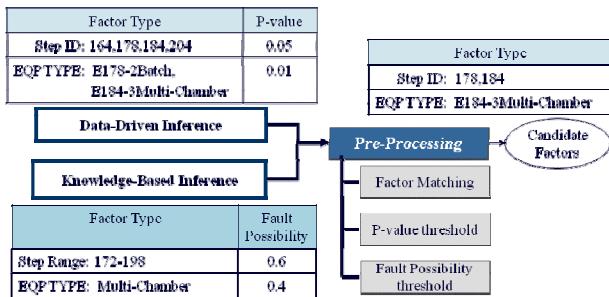


Figure 6: Generation of candidate factors for Bayesian inference

It should be noted that the granularity of the physical meanings of the factors generated by data-driven inference tools may not be the same as that generated by knowledge-based inference tools. Additional pre-process operations are needed for matching the factors generated from different type of systems.

As illustrated in Table 1, the physical meaning of the factor X generated by data-driven tools such as decision tree, ANOVA and Stepwise Regression could be (S: Step ID), (E: Equipment ID), (T: Time Period), (R: Recipe ID), and (I: In-line Parameter). On the other hand, the physical meaning of the factor X generated by knowledge-based inference tools such as Fuzzy Inference and Rough Set Theory could be (S:Step ID or a range of Step Ids), (E: Equipment type such as Batch, Single-Wafer, Multi- Chamber, etc), and (I: In-line Parameter).

Figure 6 demonstrates an example of matching the factors generated from different type of systems. Suppose the factors generated by the data-driven and knowledge-based inference tools are as followings:

- Data-Driven Inference:
S: 164, 178, 184, 204
E: E_{178-2Batch}, E_{184-3Multi-Chamber}
- Knowledge-Based Inference:
S: 172~198
E: Multi-Chamber

After the successive operations of factor matching and intersection, the final candidate factors are

- S: 178, 184
E: E_{184-3Multi-Chamber}.

Table 1: Comparisons between the factors generated by data-driven tool and knowledge-based system.

Factor: X	Data-Driven Tool	Knowledge-Based System
Inference Method	ANOVA, Decision Tree, Stepwise Regression, etc.	Fuzzy Inference, Rough Set Theory, etc.
Physical Meaning	<input type="checkbox"/> S: Step ID <input type="checkbox"/> E: Equipment ID <input type="checkbox"/> T: Time Period <input type="checkbox"/> R: Recipe ID <input type="checkbox"/> I: In-line Measurement	<input type="checkbox"/> S: Step ID or a range of Step ID <input type="checkbox"/> E: Equipment type (Batch, Single-Wafer, Multi-Chamber, etc.) <input type="checkbox"/> I: In-line Measurement
Significance	<input type="checkbox"/> Information Gain (Entropy) <input type="checkbox"/> P-Value (Statistics Significance)	Possibility (Subjective Belief)
Effect	Continuous ($Y = \beta_j X_j$)	Discrete (+H, +L, -L, -H)

4 ISSUES OF BAYESIAN INFERENCE

4.1 Prior Generation

The first stage of Bayesian Inference is to generate the prior distribution of each explanatory variable. The prior distribution is to express the engineer's subjective judgments on each explanatory variable in the probability distribution format. With the prior distribution, we can then integrate the engineer's subjective judgments with the objective evidences shown by the response data using the rigorous probability calculus. The major issues of prior distribution generation are (1) how to extract the engineer's subjective judgments and (2) how to transfer the engineer's subjective judgments into the probability distribution for-

mat. The corresponding strategies to cope with the two issues are briefly described as follows.

- (1) Use the fault-possibility values from knowledge-based inference tools to represent the engineer's subjective judgments. It is difficult to know the subjective judgments made by individual engineers under all situations. Nevertheless, it is possible to infer the common judgments made by a group of engineers with respect to various situations. The expert system with knowledge-based inference mechanism resembles a group of senior engineers in making judgments on various situations. As a result, the fault possibility of each explanatory variable generated by the expert system could be served as the engineer's subjective judgments.
- (2) Transfer the fault-possibility values generated by knowledge-based inference tools into the prior distributions of regression coefficients. Suppose there is a regression model to relate the explanatory variables to the response variable. The engineer's subjective judgment on the contribution of a specific explanatory variable to the response variable could then be specified by the prior distribution of the corresponding regression coefficient. As a result, it needs a specification of a family of prior distributions for regression coefficients, and a mapping function to transfer the fault-possibility values into the prior distributions. For example, the specification of prior distributions for the regression coefficients could be a "uniform-except-impulse-at-zero" distribution (Figure 7). That is, the regression coefficient is uniformly distributed within a specified interval except for an impulse of probability mass occurred at zero. In this way, the mapping function could then be simplified as the rule: the smaller the value of fault-possibility for an explanatory variable, the higher the impulse of probability mass occurred at zero of the corresponding regression coefficient.

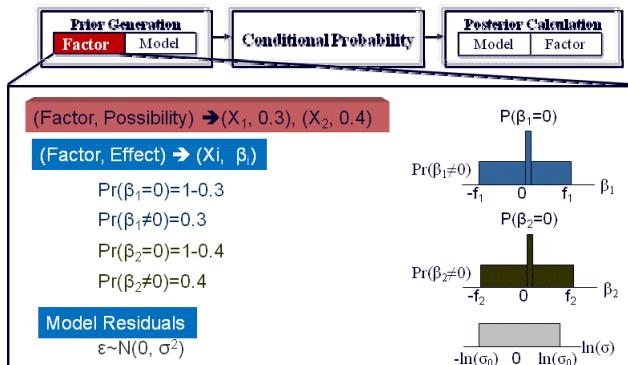


Figure 7: Factor prior distribution

To achieve better quality of the posterior distribution, a Bayesian variable selection in linear regression approach was developed by Mitchell and Beauchamp (1988). First, a set of candidate explanatory variables is selected to form a full regression model for the simultaneous explanation of response variations. Then, based on the prior distributions of individual regression coefficients, the prior model distributions $\Pr(A_m)$ will be derived for the full regression model and all the possible reduced regression models, which omit a specific subset of explanatory variables from the full regression model (Figure 8).

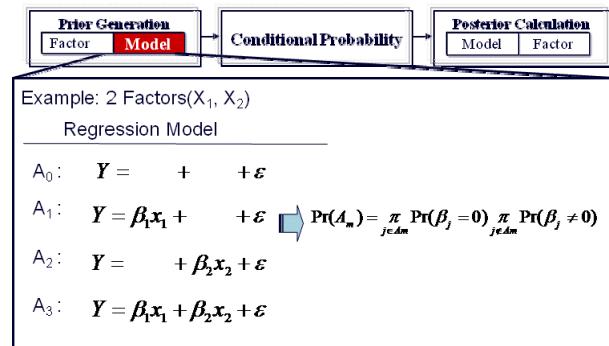


Figure 8: Model prior distribution

4.2 Conditional and Posterior Probabilities Calculation

The follow-up calculation of conditional probability is demonstrated in Figure 9 (Mitchell and Beauchamp 1988). Finally, the posterior model distributions $\Pr(A_m|y)$ and posterior probability density $p(b_i|y)$ is demonstrated in Figure 10 (Mitchell and Beauchamp 1988). Instead of the greedy selection of explanatory variable for model evaluation, this strategy considers all the possible reduced models and is expected to have high quality of the posterior distribution computation. However, the computation efficiency will be greatly reduced. To further alleviate the computational burden, an alternative technique, Gibbs Sampling, is proposed for the rapid selection of promising variables (George and McCulloch 1993).

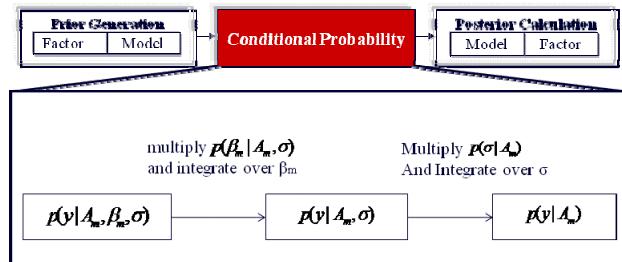


Figure 9: Conditional probability distribution calculation

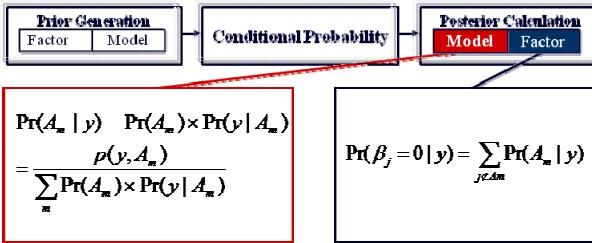


Figure 10: Posterior calculation

5 POST-PROCESSING FOR BAYESIAN INFERENCE

5.1 Intuitive Factor Ranking

To facilitate the yield-loss factor identification, an intuitive approach is to conduct the order-ranking on all suspected factors using the posterior probability. Nevertheless, there are too many assumptions and approximations on the modeling of prior probability and conditional probability, which may lead to the unreliable posterior probability. The single indicator of posterior probability cannot fully earn engineer's trust. What the engineer needs is not just the final output of posterior probability and its order ranking. For in-depth interpretations, engineers are also interested in the original information used for deriving the posterior probability, i.e. statistical p-value and fault possibility inferred by the data-driven and knowledge-based tools respectively. Figure 11 is an example to illustrate the order-ranking of suspected factors $X_1 \sim X_5$. For each suspected factor, three significance indicators for order ranking are considered:

- **Statistical p-value:** The p-value is statistically inferred by data-driven tools. The smaller the p-value the more statistical significance.
- **Fault Possibility:** A subjective belief made by engineers or knowledge-based systems. The larger the fault possibility of a factor, the more likely the factor is abnormal.
- **Posterior Probability:** The larger the posterior probability of a factor, the more likely the factor is abnormal.

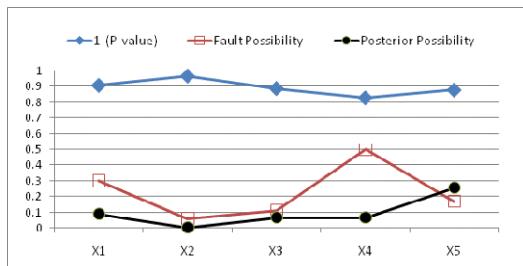


Figure 11: Intuitive factor ranking

As can be seen from Figure 11, the interpretations made by the three significance indicators are different from each other. The indicator of 1-(p-value) shows that X_2 is the most significant factor. On the other hand, the fault possibility generated by knowledge-based system points out that X_4 is most likely to be faulty. As for the Bayesian calculation, X_5 has the largest posterior probability and is inferred to be the abnormal factor. How to conduct the factor-ranking based on the combinational use of three indicators remains to be answered.

5.2 Design of Bubble Diagram with Pareto Frontier for Factor Ranking

The decision making by the combinational use of three significance indicators is somehow like the multi-objective optimization. Therefore, the concept of Pareto Frontier (Wang and Shan 2005) for multi-objective optimization is applied to the order-ranking of suspecting factors for yield diagnosis. A bubble diagram with Pareto Frontier is therefore designed to facilitate the simultaneous interpretation of three indicators

Figure 12 demonstrates a bubble diagram with Pareto Frontier, where each bubble corresponds to a suspected factor. The horizontal axis represents the (1-fault possibility) while the vertical axis represents the p-value. The smaller the (1-fault possibility) and p-value are, the more significance the factor is. As for the bubble size, it is proportional to the factor's posterior probability. A suspected factor with a larger bubble is more likely to be the faulty factor.

Without considering the posterior probability, i.e. the bubble size, the bubbles (X_1 , X_2 and X_4) are not strictly dominated by any other bubbles. Hence the bubbles (X_1 , X_2 and X_4) lie on the Pareto Frontier and should be investigated with the first priority

However, among the three significance indicators, posterior probability is the integrated and principle indicator. The other two indicators, p-value and (1-fault possibility), are supplementary indicators. Therefore, in addition to watch the Pareto Frontier, we should also consider the bubble size, i.e. posterior probability. When the bubble size is taken into consideration, we may interpret that the factor corresponding to bubble X_5 is the faulty factor, though bubble X_5 is dominated by bubble X_1 from the perspective of the other two indicators.

It should be noted that when we interpret the charts in Figure 11, the factor corresponding to the bubble X_1 in Figure 12 is very likely to be neglected. However, the bubble X_1 should be definitely highlighted, because it lies on the Pareto Frontier and is also more significant than other bubbles (X_4 and X_2) on the Pareto Frontier. The Bubble Diagram with Pareto Frontier provides engineers a more clear and objective approach to visually rank the suspected factors.

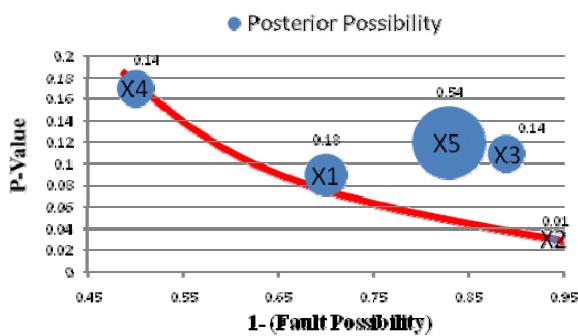


Figure 12: Pareto Frontier in Bubble Diagram

6 CONCLUSIONS

The deficiencies of data-driven techniques commonly used for yield diagnosis are characterized as the FI & MI problems. To cope with FI & MI, a preliminary study of designing a Bayesian framework to integrate both data-driven and knowledge-based inference tools in industry practices are also conducted. Specific issues of data pre-processing, Bayesian model selection, and data post-processing are also addressed, which provide directions for implementing a real system.

REFERENCES

- Chien, C. F., W. C. Wang, and J. C. Cheng. 2007. Data mining for yield enhancement in semiconductor manufacturing and an empirical study. *Expert Systems with Applications* 33:192-198.
- George, E. I., and R. E. McCulloch. 1993. Variable selection via gibbs sampling. *Journal of the American Statistical Association* 88:881-889.
- Guo, R. S. 1996. Intelligent process diagnosis based on end-of-line electrical test data. In *Proceeding of the 1996 Electronics Manufacturing Technology Symposium*, ed. C. K. Tsai, J. H. Lee, and S. C. Chang, 347-354.
- Mitchell, T. J., and J. J. Beauchamp. 1988. Bayesian variable selection in linear regression. *Journal of the American Statistical Association* 83:1023-1032.
- Wang, G. G., and S. Shan. 2005. An efficient pareto set identification approach for multi-objective optimization on black-box functions. *Journal of Mechanical Design* 127:866-874.

AUTHOR BIOGRAPHIES

CHIH-MIN FAN is an Assistant Professor in the Department of Industrial Engineering and Management at Yuan Ze University, Taiwan. He received a master's degree in

Electrical Engineering and a Ph.D. in the same subject from National Taiwan University, Taiwan. His current research and teaching interests are in statistical data analysis, knowledge representation and reasoning, engineering knowledge mining, and on-demand knowledge service with applications to engineering chain collaboration in semiconductor manufacturing industry.

YUN-PEI LU is a master student in the Department of Industrial Engineering and Management at Yuan Ze University, Taiwan. Her research interest is on the integration of data-driven and knowledge-based inference tools to enhance the quality of inference results for better decision makings, especially in the case of small samples but with large number of decision parameters .