Multivariate Outlier Modeling for Capturing Customer Returns – How Simple It Can Be

Jeff Tikkanen[1], Nik Sumikawa[2], Li-C. Wang[1], Magdy S. Abadir[2]
[1]UC-Santa Barbara, [2]Freescale Semiconductor

***Abstract*—Univariate outlier analysis has become a popular approach for improving quality. When a customer return occurs, multivariate outlier analysis extends the univariate analysis to develop a test model for preventing similar returns from happening. In this context, this work investigates the following question: How simple multivariate outlier modeling can be? The interest for answering this question are twofold: (1) to facilitate implementation of a test model in test application and (2) to ensure robustness of the methodology. In this work, we explain that based on a Gaussian assumption, a simpler covariance-based outlier analysis approach can be sufficient over a more complex density-based approach such as one-class SVM. We show that correlation among tests can be a good metric to rank potential outlier models. Based on these observations a simple outlier analysis methodology is developed and applied to effectively analyze customer returns from two automotive product lines.**

1. Introduction

Customer returns are parts that pass all tests but fail at customer site. In automotive market, the business target is to achieve zero return. When rare returns occur they are analyzed thoroughly. The outcome of the analysis usually leads to modification or addition of screens to ensure no similar occurrence in the future.

Customer returns are mostly due to latent defects. This is especially true for automotive products where a very comprehensive test flow is applied to ensure zero test escape.

In the automotive market, Part Average Testing (PAT) [2] is a common approach to screen abnormal parts based on parametric tests. There can be two types of PAT, static and dynamic, where both look to screen univariate outliers.

Suppose a part passes all the univariate outlier tests and fails at customer site. A natural extension to the outlier analysis is to look for multivariate outlier models [4]. A multivariate outlier model is constructed in a *test space* defined by multiple tests collectively. Search for a multivariate outlier model require searching for the appropriate test space. One can call this a *test space search* problem.

While prior works had proposed the idea of using multivariate outlier screening [4][11], there are several fundamental questions unanswered. First, finding an outlier model by itself does not justify the application of the model. This is because with enough tests that can be chosen to define a multivariate space, many good parts can also become "outliers."

Second, for a given customer return, there can be more than one multivariate outlier models to choose from. This leads to another fundamental question: how many multivariate outlier models are there for a given return?

Third, even though one can build a outlier model in simulation, it does not mean the model can be applied in production. For example, an SVM outlier model [9] is represented by a collection of samples. Such a model may be too complex to implement on a tester or online. Hence, a simpler model is always preferred.

To address these fundamental questions, in this work we first show that for many tests, the Gaussian assumption can be quite reasonable to characterize their distributions. Then, based on the assumption this work investigates the next question: What is the *coverage space difference* between a collection of univariate models and the multivariate model based on the same subset of tests?

We observe that this coverage space difference is larger for highly correlated tests than that for uncorrelated tests. This observation leads to a way to prioritize the outlier test spaces where test correlations are used to determine their ordering to be examined. This prioritization gives a simple strategy to tackle the test space search problem and when applied to an automotive product line, uncovers multivariate models that can be effectively applied.

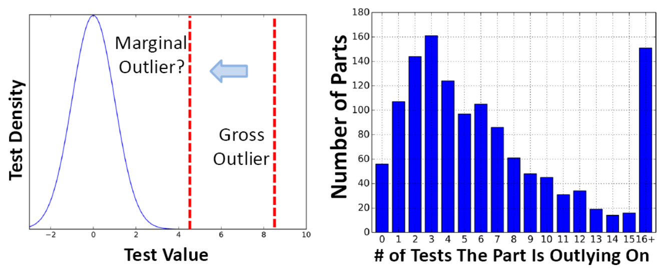
The rest of the paper is organized as follows. Section 2 explains that being an outlier does not imply being abnormal. Section 3 shows that given a return, there can be many multivariate models to consider. This motivates the development of a strategy to prioritize the outlier test spaces. Section 4 shows that for the given test data under study, most tests result in a distribution that is Gaussian. Based on the Gaussian assumption, we then explain why a simple covariance-based outlier model building technique can be used to replace a complex density-based technique like one-class SVM. Section 5 discusses the difference between applying a collection of univariate outlier models (e.g. SPAT and DPAT) and a multivariate model using the same subset of tests. Based on the observation in Section 5, section 6 suggests a simple strategy to prioritize the outlier search process based on dimensionality and test correlations, and demonstrate its effectiveness on cutomer returns from two automotive product lines. Section 7 concludes with final remarks.

2. Being outlier $\neq$ being abnormal

Outlier analysis is a form of unsupervised learning. Outlier is a relative measure. Hence, to identify an outlier, one first has to define a population set used to define the boundary of inliers. We can call this set the *base set*.

In this work, we consider wafer sort tests. Given a test distribution formed by the measured values from a base set of dies, deciding an outlier boundary can be subjective. The base set can be all dies from a wafer, multiple wafers or a lot. And typically one looks to screen out ”gross” outliers whose measured values are far from the distribution. This is illustrated in Figure 1-(a). However, deciding how far is far enough can be subjective. This decision is also impacted by the concern of yield loss. Therefore, typically one would not set the boundary close to the distribution.

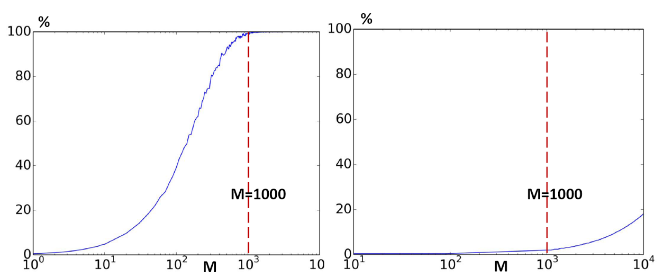
Assuming that ”gross” outliers had been screened out, when a customer return presents, one would look for screening the return as a ”marginal” outlier. However, as the outlier boundary moves closer to the distribution, many good dies may become outliers as well.



(a) Outlier is subjective (b) Good dies can be “outleirs”

Fig. 1. Many good dies can be ”outliers”

Figure 1-(b) plots an outlying property based on more than 1K good dies. The test data is collected for an airbag sensor part with 950+ parametric tests. The x-axis shows the number of tests a die in outlying on, where being an outlier is defined as being among the top five most outlying dies. As the plot shows, only less than 60 dies are not classified as an outlier (outlying in no test) based on the particular outlier definition. The rest are outlying in one or more tests, with more than 150 dies outlying in ≥ 16 tests.



(a) % of dies outlying in at least 1 test (b) % of 2 dies outlying in at least 1 same test

Fig. 2. Stat. simulation of outlying properties

Plot in Figure 1-(b) can be explained with a simple statistical simulation. Assume that there are M tests and 1K dies. Further assume that the measured values of each test follow the same Gaussian distribution. Assume the measured values of each die is randomly drawn from the Gaussian distribution. Figure 2-(a) plots the percentage of dies found to be an outlier based on at least one test. Here again, an outlier is defined to be among the top five most outlying dies in a particular test distribution.

Observe in Figure 2-(a) that, as the number of tests M grows to 1000, almost all dies are outliers. Figure 2-(a) shows that with enough tests, any die can be marginally outlying in at least one test. This simple statistical simulation confirms what we observe from the test data in Figure 1-(b).

Suppose one finds a test such that a customer return resides as one of the top five most outlying dies, Figure 1-(b) essentially demonstrates that finding such an outlying property is insufficient to justify its application. Further evidence is required to justify the model. One way can be to find additional returns that are also classified by the model as outliers. In other words, a model can be further justified if it is shared by multiple returns.

Figure 2-(b) follows the same simulation and considers all combinations of 2-die pairs. The plot shows the percentage of pairs that both dies are outliers in at least one same test. Observe that for 1000 tests, the percentage drops significantly. Therefore, when a model is shared by two returns, one has much higher confidence to apply the model than the case where the model is based on only one return. Furthermore, it is intuitive to see that this confidence grows rapidly as the model is shared by more returns.

3. The # of Models for a Return

Searching for models shared by multiple returns can begin by considering all outlier models for a return. This raises the question: How many outlier models for a return if one considers both univariate and multivariate models?

Note that a multivariate model can be based on i tests for any i less than or equal to the total number of tests. Hence, given n total tests there are $2^n - n - 1$ *test spaces* that can be used to define a multivariate outlier model. Here each test space is formed by a combination of tests.

Suppose the model building algorithm is fixed. For example we use one-class SVM [1] as suggested in [10]. Also suppose the *base set* for the outlier analysis is fixed. For example, the base set consists of all dies on the same wafer, multiple wafers or from the same lot. Further, the definition of an outlier is fixed as being among the top five outlying dies. With these aspects fixed, each test space corresponds to one outlier model. Then, Table I shows the number of possible outlier models for a given return.

In Table I, the base set consists of around 1300 dies. The ”# of tests” shows the size of the test space, i.e. using one, two or three tests. Again, there are more than 950 tests. For example, in total there are more than 142M 3-test test spaces to consider ($\frac{950\times949\times948}{1\times2\times3}$), where in 795K test spaces 1-class SVM builds a model to classify the return as one of the top 5 outlying dies among the 1300 dies.

TABLE I
THE # OF MODELS CLASSIFYING A GIVEN RETURN AS ONE OF THE TOP 5 OUTLIERS (USING 1-CLASS SVM)

# of tests	# of models	Runtime
1	11	0.25s
2	3027	59.27s
3	795128	19195.40s

4. IF TEST DISTRIBUTION IS GAUSSIAN

From the appearance, the results in Table I may motivate one to develop a search heuristic to overcome the seemingly exponential search space. For example, the heuristic can proceed by following the size of the test space, i.e. the number of tests. The heuristic explores the test spaces with i tests before exploring the spaces with $i+1$ tests. However, such a search heuristic may run out of time before determining if it can or cannot find a shared model.

Being able to answer "no model exists" is essential for an approach to be considered robust. The robustness requirement is crucial for its practical use. Otherwise, it would be difficult to know when to stop and when to resort to other means (such as failure analysis (FA)) to tackle a customer return.

4.1. Why One-Class SVM

One can overcome the challenge presented in Table I by arguing that there is no need to go beyond certain dimensionality. For example, one can show that if there is no shared model found in two-dimensional spaces, the chance of finding one in three- or higher-dimensional spaces is minimal. However, outlier modeling depends on the outlier model building algorithm. Then, one faces the question of choosing an algorithm to demonstrate the property and additionally, if using one algorithm is enough. In general, without making any assumption on the underlying distribution of the data, it is difficult to establish this property.

One-class SVM is an outlier analysis approach based on density estimation [1]. The approach makes no assumption of the underlying distribution of the data. This is in contrast to a more traditional covariance-based method where one assumes that the underlying distribution is Gaussian. In a covariance-based method, the covariance of the distribution is estimated [8]. Then, one can use *Mahalanobis Distance* to define an outlier boundary [3].

Mahalanobis distance is an adjusted euclidean distance from the mean of a distribution, where the distance is adjusted based on the covariance. It is defined as:

$$Md(x) = \sqrt{(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

where Σ^{-1} is the inverse of the covariance matrix and μ is the mean vector. In application, the mean and covariance are estimated from the data.

Assuming the data is multivariate normally distributed with d dimensions, then the Mahalanobis distance of the samples follows a Chi-Squared distribution with d degrees of freedom, denoted as $\mathcal{X}_d^2$. Let $F_d(x)$ denote its CDF. An outlier model can be built by comparing the Mahalanobis distance to $F_d^{-1}(q)$ where q is a given quantile. For example, if one desires to screen outliers at 3σ bound for a univariate Gaussian distribution, the quantile is 0.9973. The equivalence Mahalanobis distance in the d-dimensional space is $F_d^{-1}(0.9973)$. For example in the two-dimensional space $F_2^{-1}(0.9973) = 11.83$

Hence conceptually, using Mahalanobis distance to define an outlier boundary with a multivariate Gaussian distribution is similar to using the standard deviation to define a boundary with a univariate Gaussian distribution. In the univariate case, it is intuitive to use, for example a $k\sigma$ to define the boundary. In the multivariate case, we would apply the equivalent Mahalanobis distance.

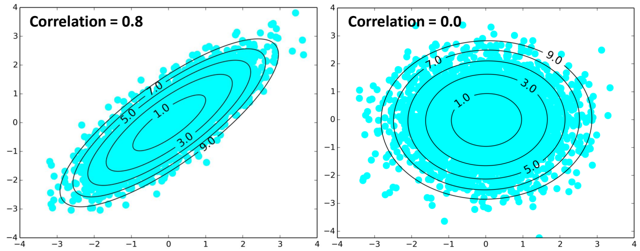


Fig. 3. Illustration of Mahalanobis Distance

Figure 3 illustrates the boundary of a Mahalanobis distance with data sampled from two-dimensional Gaussian distributions. Both distributions are based on univariate Normal distribution $\mathcal{N}(0,1)$. On the left, the two random variables are 0.8 correlated. On the right the correlation is 0. When the two random variables are correlated, observe that each Mahalanobis distance defines an oval, following the direction where the data has the most variance. In the case of correlation zero, the shape becomes a circle.

Now consider why one-class SVM may be preferred over the simpler covariance-based outlier analysis. Figure 4 illustrates their fundamental difference.

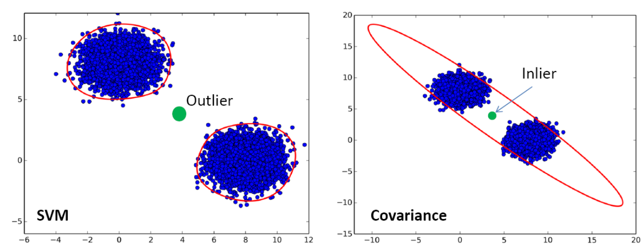


Fig. 4. 1-class SVM vs. Covariance-based model

The left of Figure 4 shows an SVM model. The right shows a covariance-based model. The underlying distribution is clearly not a single Gaussian distribution. In the SVM case, two boundaries are drawn, each is specific to a cluster of samples. The covariance-based model can only have one oval boundary. Hence, both clusters are included in the same oval. Consequently, points between the two clusters are not recognized as outliers by the covariance-based model.

Figure 4 shows that if two returns occur in between the two clusters, one-class SVM would find the shared model while the covariance-based method would miss it. This illustrates why one-class SVM is preferred and suggested in [10] - it is more powerful because it does not make any assumption about the underlying distribution of the data.

On the contrary, Figure 5 shows the SVM model and covariance-based model on a data distribution that is Gaussian. In this case, observe that the two models become very similar. This motivates the next question: Can one assume Gaussian distribution when analyzing test data?

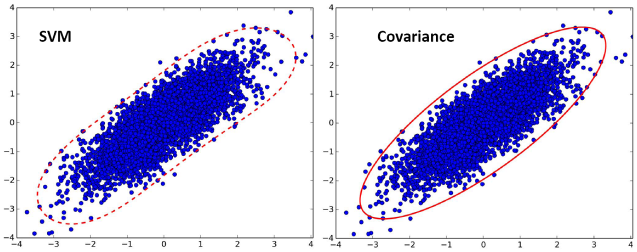


Fig. 5. Results with Gaussian-distribution data

4.2. Property of test data distribution

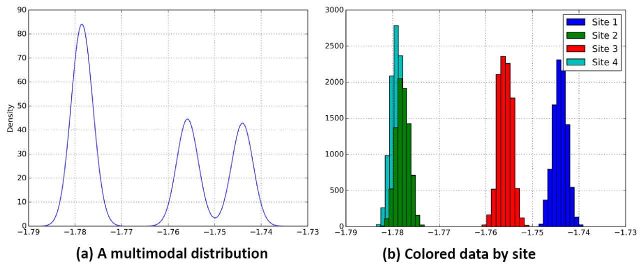


Fig. 6. Test data distribution example

Figure 6-(a) plots the test data distribution based on a particular test and all passing dies from a wafer. The data is from the same airbag sensor product line mentioned before. Observe that the distribution is not Gaussian.

Figure 6-(b), on the other hand, plots the same test data by coloring the data points with test site. There are four sites in the wafer sort test. Figure 6-(b) illustrates that the multimodal distribution observed in Figure 6-(a) is actually an artifact of tester site-to-site variation, i.e. the wafer distribution in Figure 6-(a) is a mixture of four site distributions.

What if one aligns the means of the four site distributions? Figure 7-(a) shows the result. The distribution looks Gaussian. In fact, a statistical test confirms that the distribution indeed can be assumed Gaussian with high confidence. Figure 7-(b) shows another result from another test. The distribution is also confirmed to be Gaussian by the statistical test.

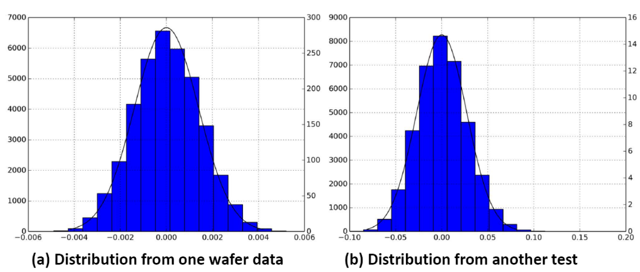


Fig. 7. Distributions after removing site variation

There are many ways to test if a given set of data points follows a certain distribution. For example, Q-Q plot [12] is one of the popular methods. Given a CDF F, the quantile for a value q is $F^{-1}(q)$. In a Q-Q plot, the quantiles from the assumed distribution (i.e. Gaussian) are plotted against the quantiles calculated based on the data for a range of q values. In the ideal case they match in every pair. One can calculate the correlation between these two sequences of quantiles to measure how well they match. For example, in the test performed for Figure 6, the Gaussian assumption is confirmed with such a correlation > 0.96.

We applied the statistical test to wafer test data from each individual test after removing the site-to-site variation. Among all tests, 93% of them are confirmed to follow a Gaussian distribution. 1.3% of them follow an exponential distribution. Others fail both types of statistical tests. The results are checked across multiple wafers from different lots.

4.3. Covariance-based modeling

If majority of the tests result in measured values that follow a Gaussian distribution, then according to Figure 5, the benefit of applying the more complex one-class SVM algorithm over the covariance-based method diminishes. This is further illustrated in Figure 8.

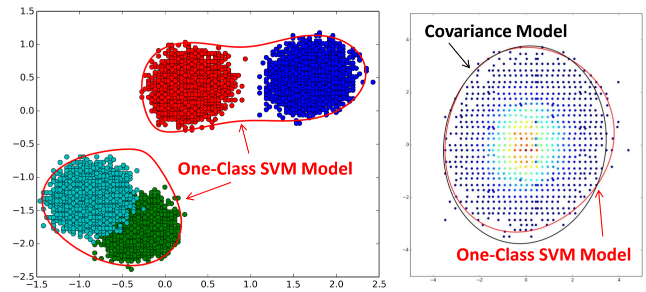


Fig. 8. Covariance modeling is enough

Figure 8 is based on the same set of dies used to plot Figure 6 by adding a second test to make it a two-dimensional plot. The left plot shows the original distribution in the two-dimensional test space with site-to-site variation. The right plot shows the distribution by aligning the means of the four clusters. In the right plot, observe that the the difference between the two models is very small.

Because the difference is rather small and because we look for models shared by at least two returns, there is no clear benefit of one method over the other. If one selects to use one method only, the chance of missing a shared model by using another method is minimal. Hence, under these assumptions the covariance-based method would be preferred because it is simpler and hence, easier to examine its properties.

We choose to use the covariance-based method also because our original goal is to find a way to prune the search space presented in Table I. The simplicity of a covariance-based model allows one to examine the space coverage difference more easily. By *space coverage difference*, we mean the region in a test space that is defined as outlier region by one model but as inlier region by another model.

5. Why multivariate outliers

What is the added value provided by multivariate outlier analysis? To see this, we analyze the space coverage difference between a multivariate model and a collection of univariate models based on the same subset of tests.

Figure 9 illustrates the space coverage difference in a two-dimensional test space. The x-axis and y-axis show the

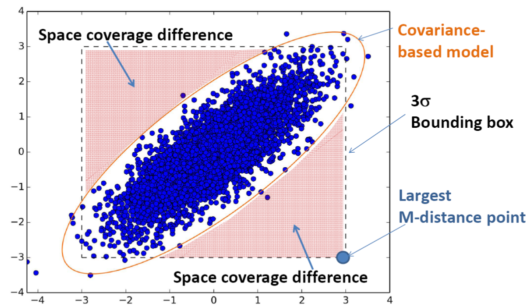


Fig. 9. Space coverage difference between a 2-dimensional covariance-based model and two univariate models using the same tests

measured values from two tests. This is an artificial example for illustration purpose.

In this artificial example, measured values from each test follows the same Gaussian distribution $\mathcal{N}(0,1)$. Hence, when combining the data from the two tests, the result is a multivariate Gaussian distribution. Measured values from the two tests are assumed to be 0.8 correlated.

For each test, the test limits to define a univariate outlier is set at $\pm 3\sigma$. The two univariate outlier models prescribes a "3σ bounding box" where dies outside the box are outliers.

The covariance-based analysis using the equivalent 3σ Mahalanobis distance gives a model of an oval shape as discussed before. Hence, the space coverage difference between the covariance-based model and the two univariate model is the shaded areas between the bounding box and the oval.

Dies inside the shaded areas are classified by the covariance-based model as outliers, but these dies would have been missed by the univariate outlier analysis. Hence, the space coverage difference is the test space where the covariance-based model provides unique coverage.

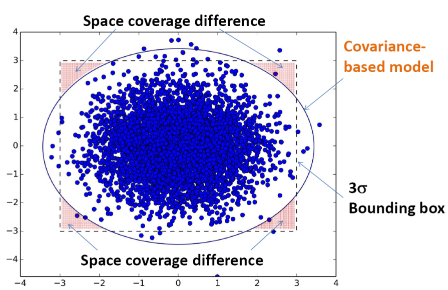


Fig. 10. Test correlation $= 0$

5.1. *Correlation and test space coverage*

Figure 10 re-plots Figure 9 by changing the assumption that two tests are 0.8 correlated to no correlation. Observe that the space coverage difference shrinks, i.e. the unique coverage by a covariance-based multivariate model shrinks.

Figure 11 illustrates the relationship between the correlation of the two tests and the unique coverage provided by the covariance-based multivariate model. On the left, the boundary of each covariance-based model is plot against the corresponding correlation. The correlation numbers are shown as 0 to 100 (%) and colored differently. On the right, the x-axis shows the correlation and y-axis shows the unique coverage measured as a percentage of area of the 3σ bounding box. We see that

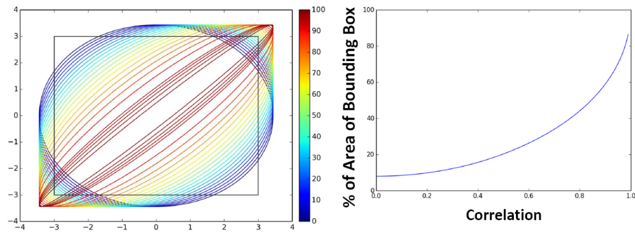


Fig. 11. Correlation vs. test space coverage

as the correlation approaches 1, the unique coverage reaches above 90% of the bounding box.

Figure 11 suggests that given all the covariance-based models of the same dimensionality, one would prefer the models with tests that have a higher correlation than the model with tests that have a lower correlation. However, it also shows that even with tests of no correlation, there is still missing coverage space by the bounding box, which can be uniquely covered by the covariance model.

6. PRIORITZING THE SEARCH

Given n tests, there can be $2^n - n - 1$ multivariate models. Let S_i be the collection of all multivariate models of the same dimensionality i, for $i = 2, \ldots, n$. Suppose one applies all model in S_2. What is the coverage space missed by S_2 and covered by $S_3 \cup \cdots \cup S_{n-1}$?

TABLE II
COVERAGE MISS (MEASURED AS %) AFTER APPLYING ALL MODELS UP TO DIMENSIONALITY i

dimensionality	2	3	4	5	6
corr=0.9	0.04	0.03	0.00	0.00	0.00
corr=0.0	5.55	0.46	0.01	0.01	0.01

Extending Figures 9 and 10, Table II illustrates this coverage miss for $n = 7$. For example, with dimensionality 3, we assume that all models in $S_2 \cup S_3$ are applied. Then, we estimate the unique coverage contribution from $S_4 \cup \cdots \cup S_6$.

The missing coverage space is estimated through Monte Carlo simulation where 1M sample points are randomly drawn inside the 3σ bounding box. Then, if the point falls outside a covariance model, it is covered. Otherwise, it is missed. Each model follows the equivalent Mahalanobis distance of 3σ bound in the univariate case. Each test again is assumed to follow the Gaussian distribution $\mathcal{N}(0,1)$. Two cases are considered: correlation=0.9 and 0.0.

The base of the % number is the total sample points that can be captured by all models, i.e. $S_2 \cup \cdots \cup S_6$. As Table II shows, after we apply all models in S_2, there is little missing coverage space. Table II suggests that after S_2 one may ignore models in higher dimensionality.

6.1. *Experimental result*

The analysis above can be summarized into three points: (1) If one has to choose between two models of the same dimensionality, the model using tests of higher correlation should be prefered. (2) If one has to choose between two models of different dimenionalities, the model with lower dimensionality should be preferred. (3) After one considers all models up to i dimensionality for $i \geq 2$, the unique contribution from rest of the models in higher dimensionality is diminishing rapidly as i increases.

Following these three points, in practice we implement a simple search strategy by exhaustively considering all models of two tests. Then, among those models we rank them based on test correlation. We applied this strategy to the airbag sensor product line. Below shows result on 9 customer returns each on a different wafer.

TABLE III
OF OUTLIER MODELS IN 1- AND 2-DIMENSIONAL TEST SPACE

Return	1	2	3	4	5	6	7	8	9
dim= 1	11	6	19	23	7	12	41	22	9
dim= 2	2988	1639	8622	11892	1618	3222	18349	8253	2110

Table III shows the number of covariance-based models that classify each return as an outlier. A similar result for return 1 was shown before in Table I. Notice there is small difference between SVM and covariance-based models in the 2-dimensional case. Again, an outlier is among the top five most outlying dies.

In the univariate case, return 5 and return 7 have shared models. However, there is no shared model between other pairs. Therefore, we follow the search strategy to look for shared models in the two-dimensional test spaces.

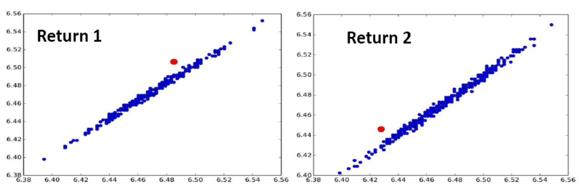


Fig. 12. Two returns projected as Mahalanobis distance based outliers

Figure 12 shows the test space that projects the two returns as outliers classified by the covariance-based models. Notice that the two tests are highly correlated. Note that there can be other shared models. This is the test space where the two tests have the highest correlation.

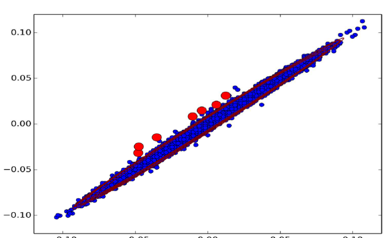


Fig. 13. Test space shared by 7 returns

The same test space is actually shared by 7 returns. Figure 13 shows the overlap of the test data from the 7 wafers and 7 returns on the same test space, i.e. taking plots such as Figure 12 and overlap them. Note that the covariance-based model is applied to the 7 wafer individually. Hence, individually each return is more outlying as shown in Figure 12 than that shown in Figure 13. On each wafer, the return is classified as one of the top five outliers.

To further demonstrate the validity of the strategy, customer returns from a 2nd product line was analyzed. The 2nd line

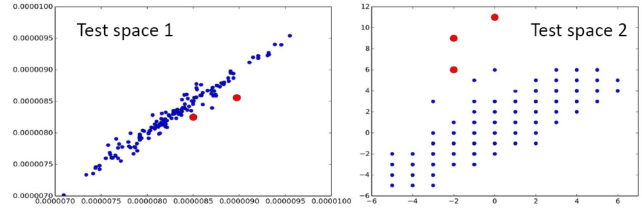


Fig. 14. Test spaces project multiple returns as outliers - 2nd product line

is an automotive SoC product with 1K+ parametric tests. Figure 14 shows two test spaces each found to be shared by multiple returns.

7. Final remarks

The use of Mahalanobis distance for screening outliers in test is not new. The authors in [5][6] had proposed using Mahalanobis distance in correlated test space to screen outliers for analog parts. Using test correlation as an indicator to look for multivariate outliers is also not new. The author [7] was among the first to suggest that, and proposed using Principal Component Analysis (PCA) to expose multivariate outliers for screening burn-in fails.

This work solves a different problem. This work shows that finding an outlier model for a known fail is relatively easy. The next challenge is to find a model shared by multiple fails. By assuming that most of the tests follow a Gaussian distribution after removing site-to-site variations, this work shows that exhaustively considering simple covariance based models in two-dimensional test spaces is a sufficient strategy. Hence, the contribution of the work is not in providing a new approach. Rather it explains when and why a simple approach such as covariance based modeling is enough.

References

[1] e. a. B. Scholkopf. Estimating the support of a high-dimensional distribution. *Neural Computation*, 2001.
[2] A. E. Council. Guidelines for part average testing. *http://www.aecouncil.com/AECDocuments.html*, 2012.
[3] D. De Maesschalck, Roy; Jouan-Rimbaud and D. L. Massart. The mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50, 2000.
[4] D. e. a. Drmanac. Multidimensional parametric test set optimization of wafer probe data for predicting in field failures and setting tighter test limits. In *DATE*, 2011.
[5] S. Krishnan and H. Kerkhoff. A robust metric for screening outliers from analogue product manufacturing tests responses. In *European Test Symposium*, 2011.
[6] S. Krishnan and H. Kerkhoff. Exploiting multiple mahalanobis distance metrics to screen outliers from analog product manufacturing test responses. *Design Test, IEEE*, 30(3):18–24, 2013.
[7] P. O'Neill. Production multivariate outlier detection using principal components. In *Test Conference, 2008. ITC 2008. IEEE International*, pages 1–10, 2008.
[8] P. Rousseeuw and K. Van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3), 1999.
[9] B. Scholkopf and A. Smola. Learning with kernels: Support vector machines, regularization, optimization,and beyond. 2001.
[10] N. e. a. Sumikawa. Forward prediction based on wafer sort data; a case study. In *IEEE ITC*, pages 1–10, 2011.
[11] N. e. a. Sumikawa. Understanding customer returns from a test perspective. In *VLSI Test Symposium (VTS), 2011 IEEE 29th*, pages 2–7, 2011.
[12] R. Wilk, M.B.; Gnanadesikan. Probability plotting methods for the analysis of data. *Biometrika*, 55(1), 1968.