EUROPE DISCOVERY SUMMIT

Root cause searching on a yield loss in automotive industry, by multivariate analysis and modeling (1/6)

Corinne Bergès, Shilu Zhang, Jim Bird, Ivan Belen, Chris Smith _ NXP Semiconductors

Abstract

- In the context of semiconductor manufacturing industry for automotive, the yield is monitored at wafer level, as a KPI in term of cost, of course, but also in term of quality as quality vs yield link is proved.
- Here, a yield loss is observed at electrical die-test step at room temperature, and in particular for a certain bin which is fitting with a group of tests of a specific component function. If a unit probing (UP) test, which the yield loss is observed for, highlights the failing dies, class probing (CP) tests the reticles that are structures built between the dies and that monitor the manufacturing steps, before the dies were functional and may be tested. So, the root cause searching will target to correlate the yield loss observed at UP step with the CP tests, in order to understand what the failing manufacturing step is.
- Correlation analysis, multivariate analysis and modeling are implemented on UP and CP data for the failing lots, using JMP PRO, and their results were fitting with some good clues for the device engineers to design the corrective actions.

Overall goals and benefits

Overall goals:

- Building reusable, automated, Machine Learning (ML) analysis tool for Failure Analysis (FA) /Device engineering Root Cause Analysis (RCA) of Class Probe to Unit Probe correlations
  - Module 1: Yield loss observed per UP bin vs CP Test
  - Module 2: Yield loss observed per UP test vs CP Test
  - Yield loss observed per FT bin vs CP Test – Hold Pending FT in Hadoop
  - Yield loss observed per FT test vs CP Test – Hold Pending FT in Hadoop
- Run repeated real-world use cases with FA and Device Eng. (a dozen), to find CP/UP correlations, and progress the tool development

Benefits:

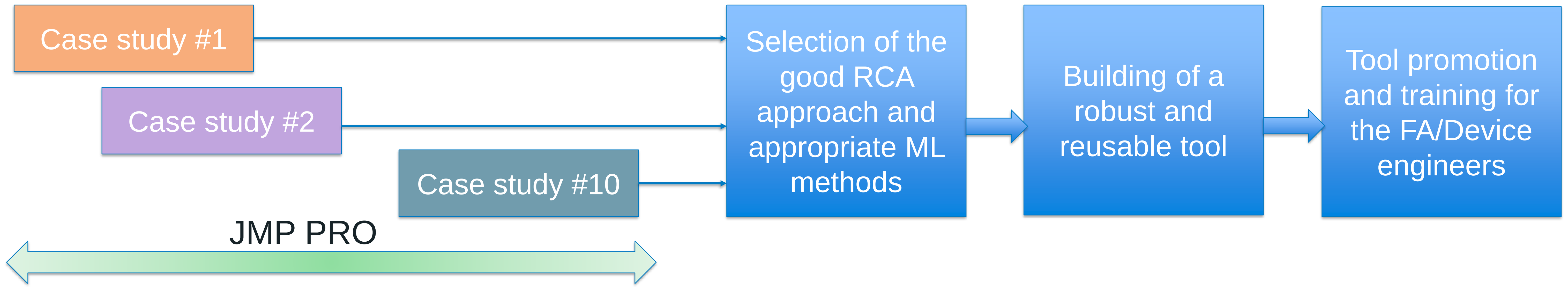
- Provide repeatable tool/process for FA and Device Engineers to easily use ML to reach RCA for CP/UP correlation
- Save FA and Device Engineers significant time during RCA

Acronyms and definitions

- Class Probe (CP): test data during manufacturing at reticle level
- Unit Probe (UP): test data at die level, on the dies still on the wafer, not yet packaged
- Final test (FT): test data at the manufacturing and assembly steps, on the packaged dies (final step before shipping to customers)
- Reticles: specific structures built between the dies: reticle test monitors manufacturing process
- A bin is fitting with a group of tests that test a specific function for the semiconductor component
- Failure Analysis (FA)
- Root Cause Analysis (RCA)
- CP, UP and FT data storage architecture is cloud-based (Hadoop, AWS)
- Hadoop, Amazon Web Services (AWS): software framework for distributed storage and processing of big data

Project plan

- Different case studies of yield losses, faced by Device Engineers, allowed to design the correct ML approach to be implemented in such types of RCA.
- In this first step, JMP appears as an interesting tool to quickly run some analysis and to obtain consistent results.



Case study

- For a specific automotive semiconductor, yield loss is observed abnormally high during a time period
- UP and CP data for 197 lots of dies are collected during this period, typically stored in Hadoop and here, extracted from Hadoop for a RCA analysis with JMP PRO
- These 197 lots are fitting with:
  - 4714 wafers (typical: 25 wafers per lot)
  - 481 CP tests
  - about 2000 UP tests (fitting with 112 bins)
- For each of the wafers, and for each of the 481 CP tests, mean of the CP test values is computed
- Yield is computed for each of the 112 bins, but Device Engineers are particularly interested by yield loss on the bin 019
- Input data file:
  - 4714 rows (1 row per wafer)
  - 481 colums of means for each CP tests + 112 columns of yield per bin

Root cause searching on a yield loss in automotive industry, by multivariate analysis and modeling (2/6)

EUROPE DISCOVERY SUMMIT

Input data

- ‘CP_UP_bin_RoomTest_241123.jmp’:
  - 4714 rows: 1 row per wafer, identified by (lot number, wafer number)
  - 481 tests in 481 columns: one cell is the mean of the test values for the dies in this wafer
  - 112 bins in 112 columns: one cell is the yield observed for this wafer and for this bin

Missing Data Pattern 241123 - JMP Pro

File Edit Tables Rows Cols DOE Analyze Graph Tools Add-I

Missing Data Pattern ...
- Source
- Treemap
- Cell Plot

	Count	Number of columns missing
1	588	0
2	4126	97

Missing data in ‘CP_UP_bin_RoomTest_241123.jmp’: 4126 data in 97 columns

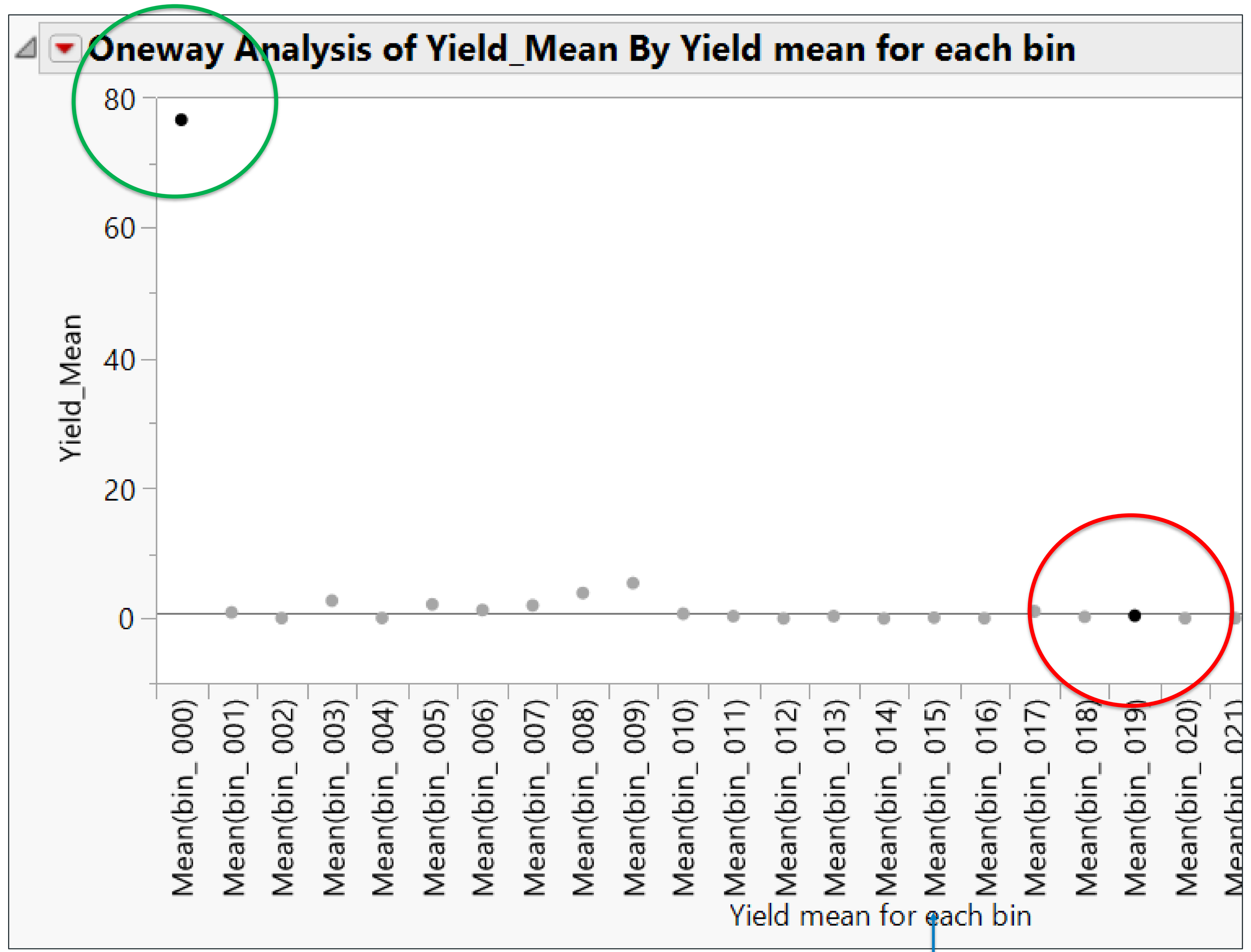
	lot_id	wafer_id	test_ 001	test_ 143	test_ 045	test_ 046	test_ 047	test_ 048	bin_ 005	bin_ 006	bin_ 007
	lot_5	5	15,2	3671	2,26p	112	19,1	28,8	8,31	9,59	23,6
	lot_8	11									
	lot_10	16									
	lot_12	17									
	lot_15	18									
	192 ...	20 others	0,95	2836	1,98p	101	14,8	20,3	0	0	0
1	lot_1	1	2,54855	3163,306	2,179787e-12	105,4116	16,37258	26,35902	0,5333333333	0,5333333333	1,3333333333
2	lot_1	10	3,110033	3182,686	2,201407e-12	103,9803	16,74158	25,26464	1,5915119363	1,8567639257	1,3262599469
3	lot_1	11	2,893351	3163,306	2,205186e-12	104,8811	16,31783	24,35934	5,497382199	1,832460733	1,3089005236
4	lot_1	13	2,843494	3167,313	2,185565e-12	107,2473	15,9473	25,22914	4,7222222222	0,8333333333	1,1111111111
5	lot_1	14	2,540263	3143,666	2,185652e-12	104,2144	16,03639	24,65724	2,3746701847	1,8469656992	1,3192612137
6	lot_1	15	2,669372	3166,812	2,18363e-12	103,8537	16,20385	24,75421	2,3316062176	1,0362694301	1,0362694301
7	lot_1	16	2,593899	3162,055	2,173838e-12	103,6571	16,2733	28,10425	3,5714285714	1,7857142857	4,0816326531
8	lot_1	17	3,054741	3168,066	2,182467e-12	104,7942	17,26602	27,26385	0	0,5376344086	1,3440860215
9	lot_1	18	3,571173	3155,072	2,158374e-12	104,0627	15,87115	25,03522	1,0610079576	0,2652519894	0,5305039788
10	lot_1	19	3,311258	3141,69	2,214452e-12	103,6247	16,11546	26,70732	1,2987012987	1,2987012987	2,3376623377
11	lot_1	2	2,406855	3169,572	2,180618e-12	104,4878	16,70897	24,67067	1,6085790885	0,2680965147	1,072386059
12	lot_1	20	3,196318	3169,07	2,188821e-12	106,2381	17,49244	24,87284	4,1899441341	1,3966480447	0,8379888268

‘CP_UP_bin_RoomTest_241123.jmp’

Frequencies

Level	Count
lot_1	24
lot_2	20
lot_3	11
lot_4	18
lot_5	25
lot_6	23
lot_7	12
lot_8	25
lot_9	22
lot_10	25
N Missing	0
197 Levels	

Number of lots



Yield mean observed per bin for all the wafers: bin_019 is the bin of interest

- Note: Bin_ 000 = passing dies → yield mean = 77%

Root cause searching on a yield loss in automotive industry, by multivariate analysis and modeling (3/6)

Correlation analysis

Correlations

- The table is a matrix of correlation coefficients (estimated by row-wise method) that summarizes the strength of the linear relationships between each pair of response (Y) variables

Multivariate

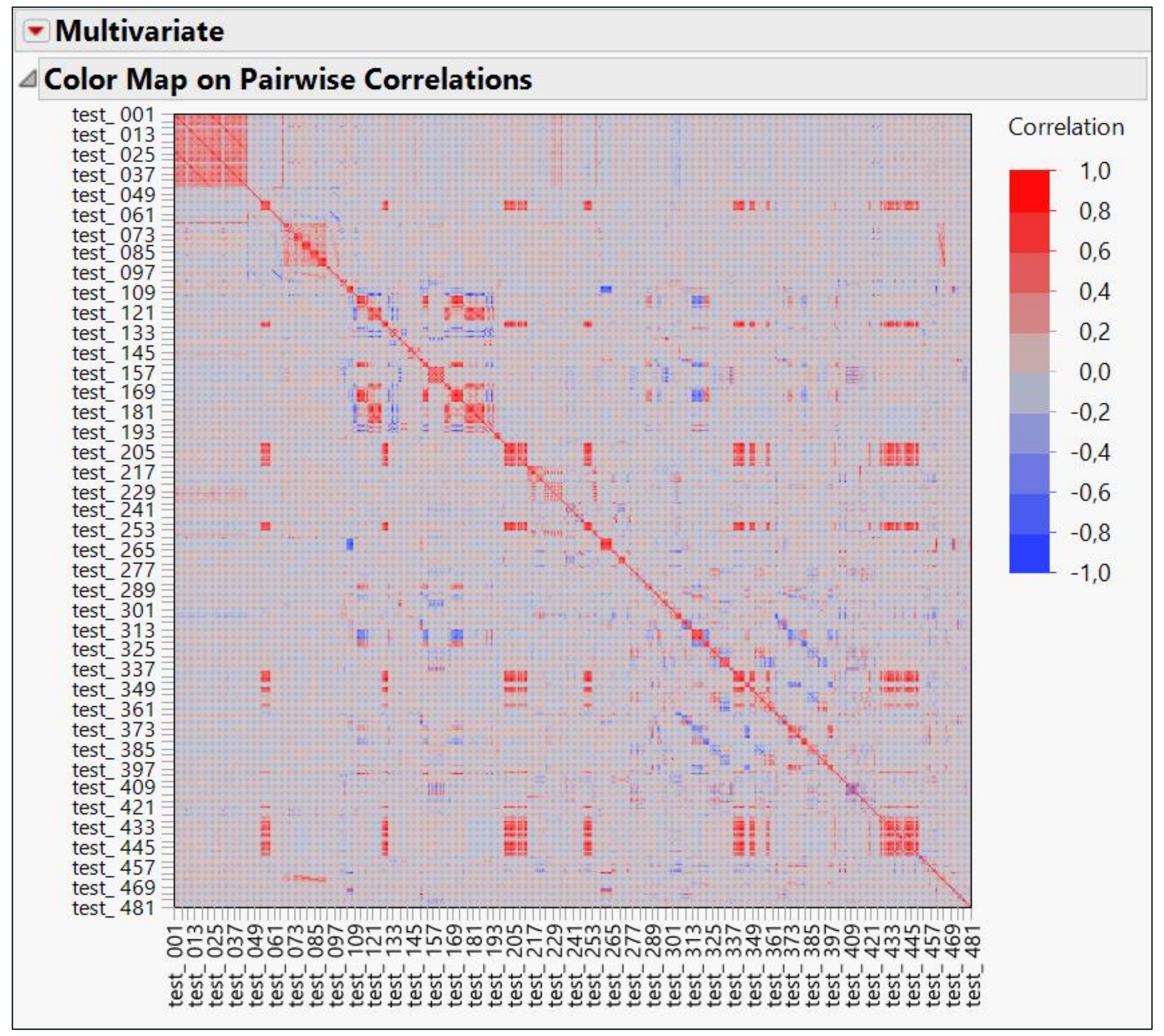
Correlations

	test_ 001	test_ 002	test_ 003	test_ 004	test_ 005	test_ 0
test_ 001	1,0000	0,4627	0,5511	0,4525	0,5207	0,40
test_ 002	0,4627	1,0000	0,4913	0,4955	0,5026	0,49
test_ 003	0,5511	0,4913	1,0000	0,4040	0,5773	0,43
test_ 004	0,4525	0,4955	0,4040	1,0000	0,4952	0,54
test_ 005	0,5207	0,5026	0,5773	0,4952	1,0000	0,51
test_ 006	0,4048	0,4987	0,4335	0,5463	0,5181	1,00
test_ 007	0,4983	0,4588	0,4498	0,4353	0,5278	0,41
test_ 008	-0,0208	-0,0239	-0,0204	-0,0183	-0,0352	-0,02
test_ 009	0,4281	0,4025	0,4228	0,3996	0,5202	0,40
test_ 010	0,3742	0,3872	0,3690	0,4396	0,4453	0,47
test_ 011	0,4682	0,4904	0,4726	0,4427	0,5488	0,41
test_ 012	0,3321	0,4411	0,3493	0,4449	0,4631	0,46
test_ 013	0,4663	0,4139	0,4236	0,4285	0,5158	0,44
test_ 014	0,3643	0,4647	0,3244	0,4601	0,4475	0,46

Correlations

Pairwise Correlations

- The Pairwise Correlations table, which lists the Pearson product-moment correlations for each pair of Y variables. The correlations are calculated by the pairwise deletion method. The count values differ if any pair has a missing value for either variable. The Pairwise Correlations report also shows significance probabilities and compares the correlations in a bar chart. All results are based on the pairwise method.



Color Map and table of the pairwise correlations

Multivariate

Pairwise Correlations

Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob
test_ 241	test_ 181	-0,2218	588	-0,2974	-0,1436	<,0001*
test_ 237	test_ 181	-0,2218	588	-0,2974	-0,1436	<,0001*
test_ 420	test_ 292	-0,0791	4714	-0,1074	-0,0506	<,0001*
test_ 373	test_ 049	0,0791	4714	0,0506	0,1074	<,0001*
test_ 357	test_ 249	0,0791	4714	0,0506	0,1074	<,0001*
test_ 338	test_ 234	-0,0791	4714	-0,1074	-0,0506	<,0001*
test_ 423	test_ 162	-0,2218	588	-0,2973	-0,1435	<,0001*
test_ 358	test_ 152	0,0791	4714	0,0506	0,1074	<,0001*
test_ 266	test_ 232	0,0790	4714	0,0506	0,1074	<,0001*
test_ 358	test_ 193	-0,0790	4714	-0,1074	-0,0506	<,0001*
test_ 414	test_ 379	0,0790	4714	0,0506	0,1074	<,0001*
test_ 294	test_ 082	-0,0790	4714	-0,1073	-0,0506	<,0001*
test_ 394	test_ 032	-0,0790	4714	-0,1073	-0,0506	<,0001*
test_ 388	test_ 369	-0,0790	4714	-0,1073	-0,0506	<,0001*
test_ 424	test_ 085	0,0790	4714	0,0506	0,1073	<,0001*
test_ 401	test_ 329	-0,0790	4714	-0,1073	-0,0506	<,0001*
test_ 411	test_ 394	-0,0790	4714	-0,1073	-0,0506	<,0001*
test_ 417	test_ 248	0,0790	4714	0,0506	0,1073	<,0001*
test_ 098	test_ 041	0,0790	4714	0,0506	0,1073	<,0001*
test_ 458	test_ 127	-0,2217	588	-0,2973	-0,1434	<,0001*
test_ 411	test_ 085	0,0790	4714	0,0506	0,1073	<,0001*
test_ 413	test_ 384	-0,0790	4714	-0,1073	-0,0506	<,0001*
test_ 471	test_ 083	-0,0790	4714	-0,1073	-0,0506	<,0001*

Statistical Details for the Pearson Product-Moment Correlation

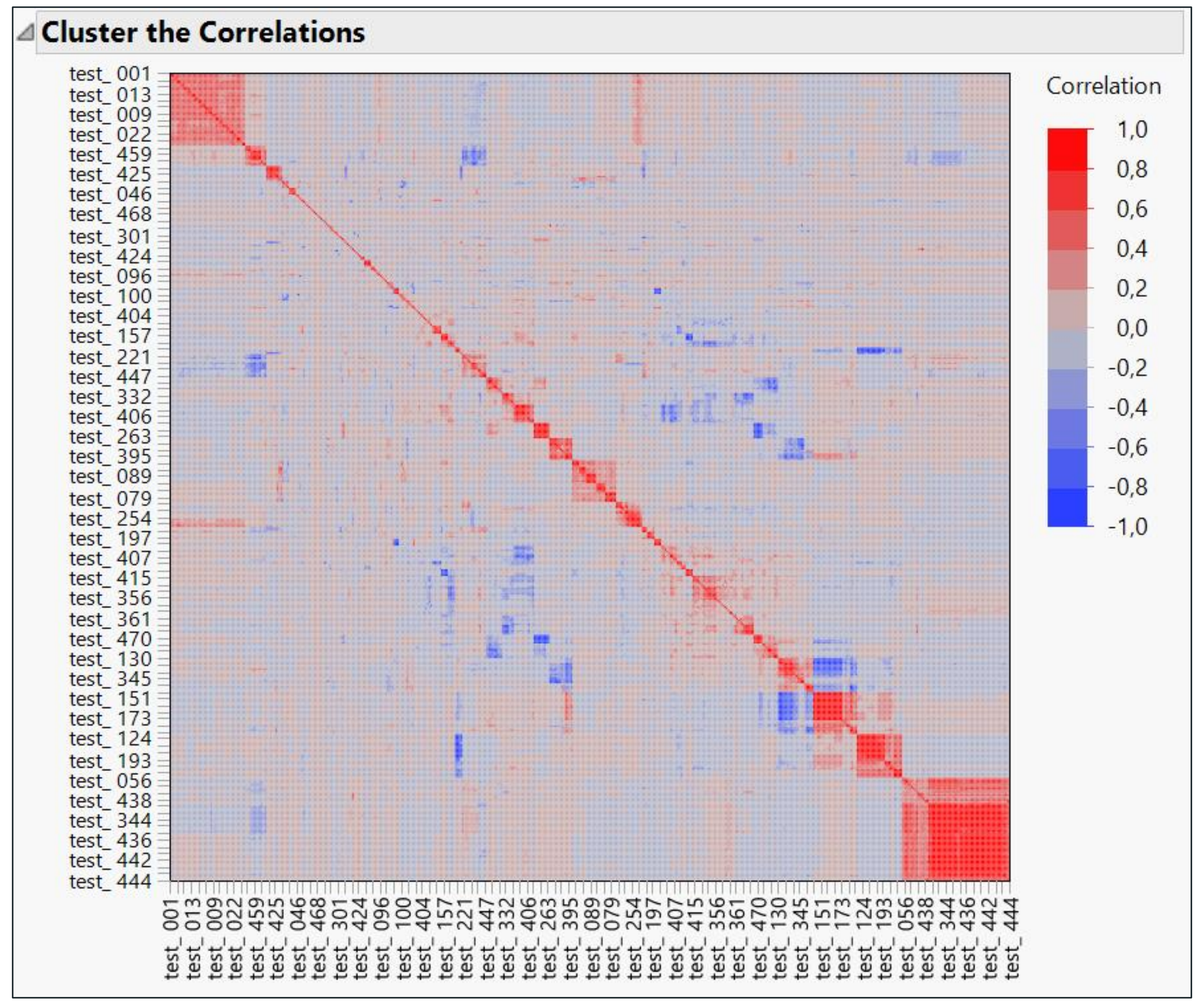
In the Multivariate platform, the Pearson product-moment correlation coefficient measures the strength of the linear relationship between two variables. For response variables *X* and *Y*, it is denoted as *r* and computed as follows:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}$$

If there is an exact linear relationship between two variables, the correlation is 1 or –1, depending on whether the variables are positively or negatively related. If there is no linear relationship, the correlation tends toward zero.

Cluster the Correlations

- Produces a cell plot that clusters together similar variables. The correlations are the same as for Color Map on Correlations, but the positioning of the variables might be different.



Correlation clusters

Comments:

- This correlation analysis and the management of the similarities are important to increase modeling efficiency.
- In this case study, this analysis allows to manage the different versions of the test programs. Indeed, a test is identified by a test number and a test name. From a test program version to the next one, it happens that the test names are modified, while corresponding with the same tests. So, the data analyst will work on new test names that are resulting from concatenating test numbers with test names. This correlation analysis allows to identify the same tests, whatever the changes that may have been performed on the test numbers and/or the test names.
- The next step is to perform clustering and to select only one test per cluster, one cluster merging the similar tests.

EUROPE DISCOVERY SUMMIT

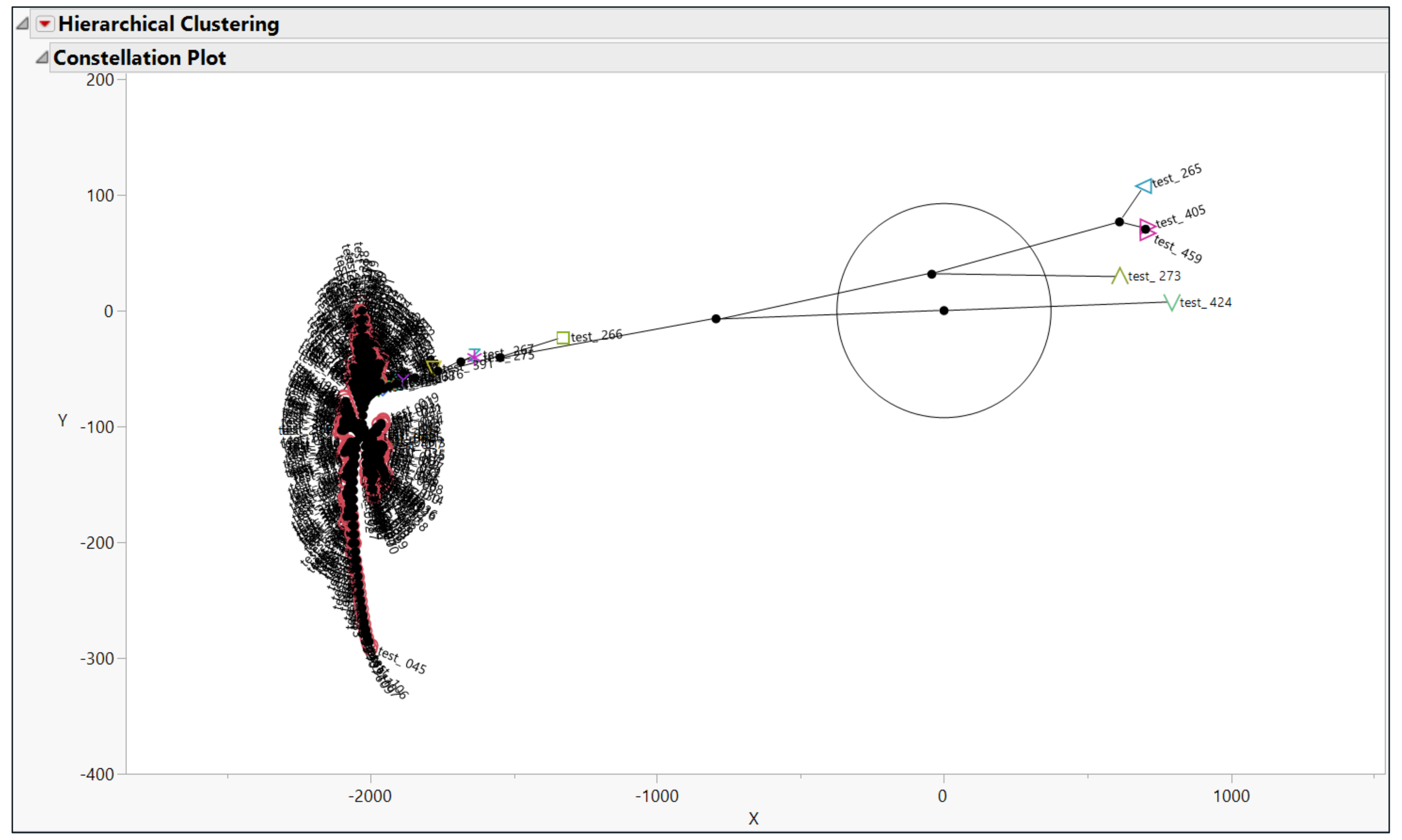
Root cause searching on a yield loss in automotive industry, by multivariate analysis and modeling (4/6)

Hierarchical clustering

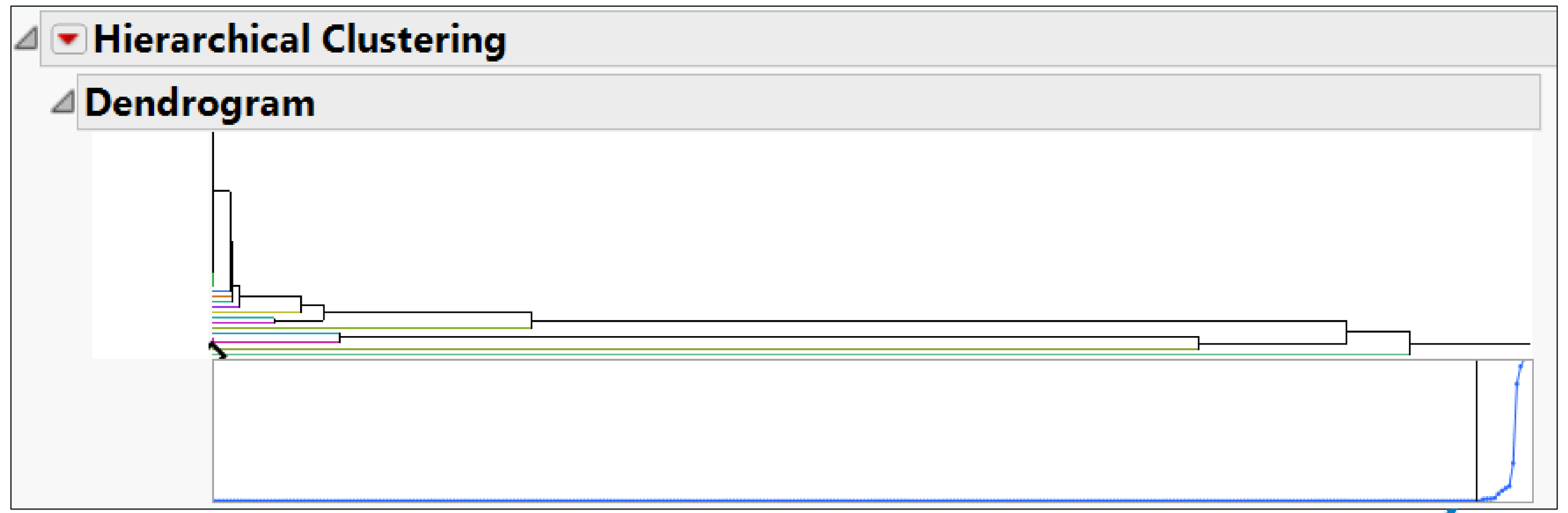
Hierarchical Clustering with Ward method

- In Ward's minimum variance method, the distance between two clusters is the ANOVA sum of squares between the two clusters summed over all the variables. At each generation, the within-cluster sum of squares is minimized over all partitions obtainable by merging two clusters from the previous generation. The sums of squares are easier to interpret when they are divided by the total sum of squares to give the proportions of variance (squared semipartial correlations).
- Ward's method joins clusters to maximize the likelihood at each level of the hierarchy under the assumptions of multivariate normal mixtures, spherical covariance matrices, and equal sampling probabilities.

Ward's method tends to join clusters with a small number of observations and is strongly biased toward producing clusters with approximately the same number of observations. It is also very sensitive to outliers.



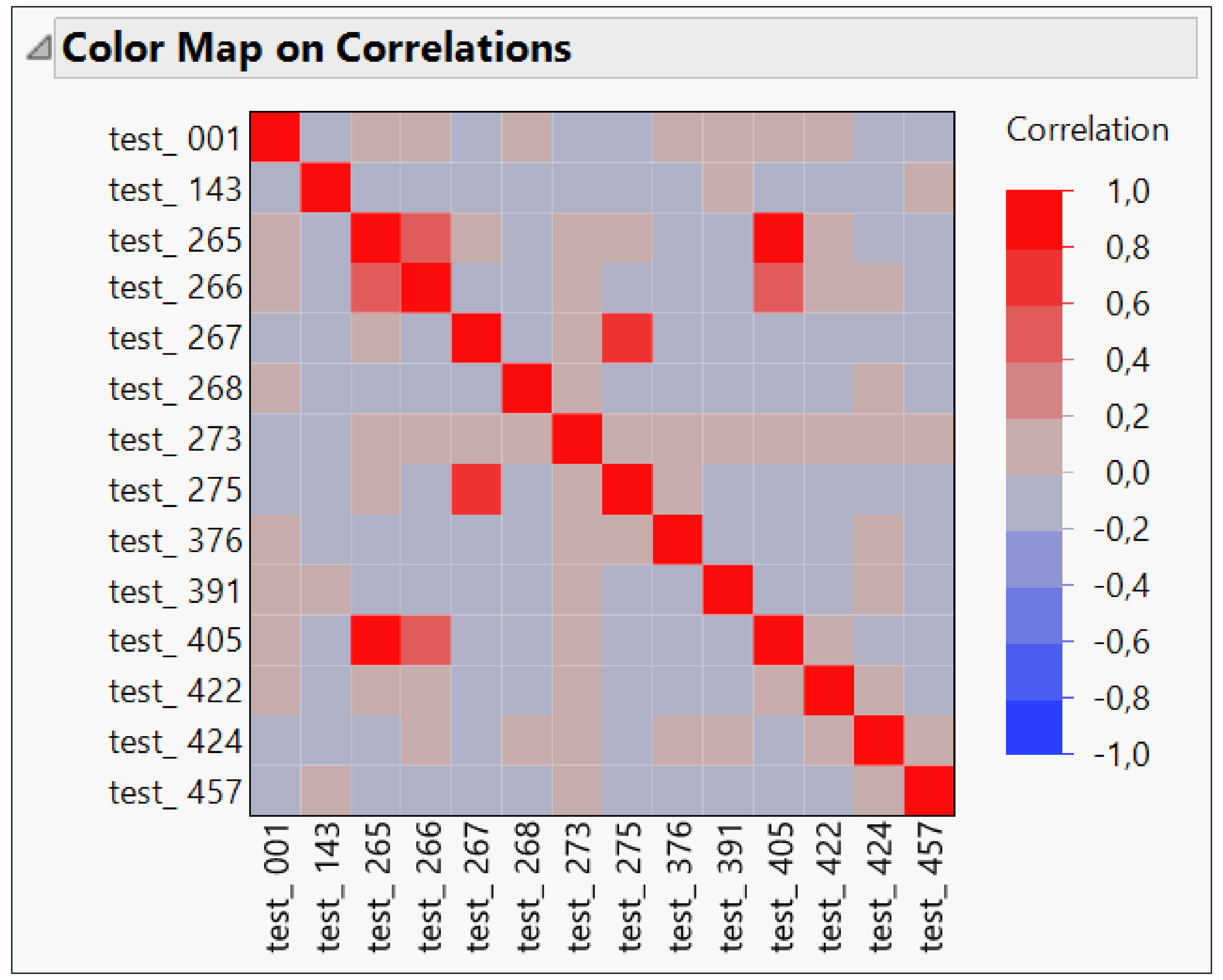
Constellation plot



Input data file: 'Transpose241123.jmp'

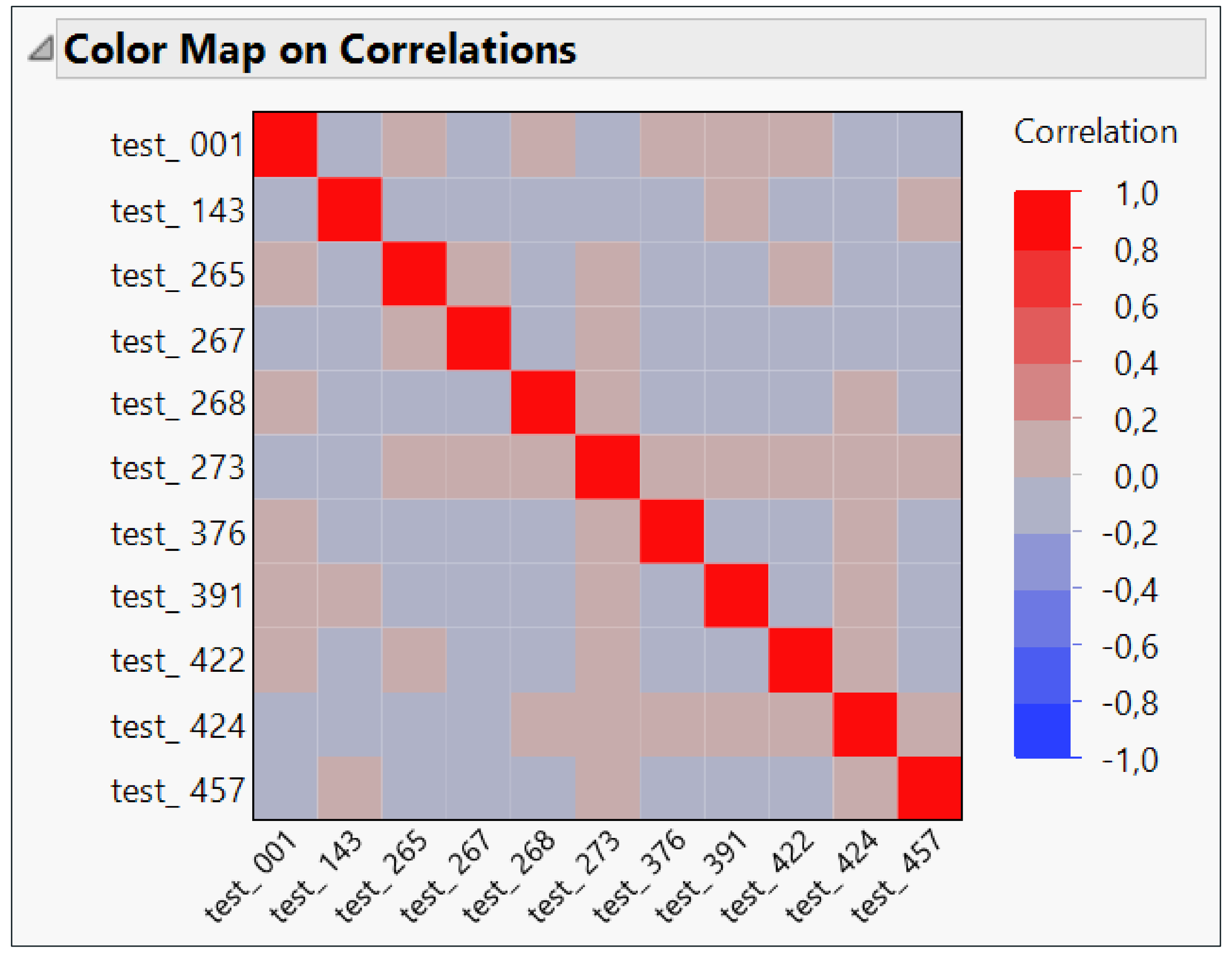
The dendrogram suggests 14 clusters

Cluster_id	Tests_used_in_modeling
1	test_001
2	test_143
11	test_265
10	test_266
8	test_267
5	test_268
13	test_273
9	test_275
6	test_376
7	test_391
12	test_405
3	test_422
14	test_424
4	test_457



A new clustering is performed to remove the last correlations: 11 clusters will be the final number of clusters selected

Cluster_id	Tests_used_in_modeling
1	test_001
2	test_143
11	test_265
8	test_267
5	test_268
13	test_273
6	test_376
7	test_391
3	test_422
14	test_424
4	test_457



EUROPE DISCOVERY SUMMIT

Root cause searching on a yield loss in automotive industry, by multivariate analysis and modeling (5/6)

Case study purpose

- In this type of analysis on a yield loss issue, and in this specific case study, the goal is to be able to identify the CP tests that could have failed during this time period, and these CP tests will highlight the failing manufacturing steps.
- Statistically, this means the list of the most contributor CP test is targeted.
- In the extent that te list of failing lots is exhaustive, interest for significancy of the obtained models is relative. The models will not be deployed elsewhere.

Measures of Fit for bin_ 019

Predictor	Creator	.2 .4 .6 .8	RSquare	RASE	AAE	Freq
bin_ 019 Predictor	Boosted Tree		0,7961	0,7193	0,3295	4714
bin_ 019 Predictor_1	Partition		0,0074	1,5871	0,5984	4714
bin_ 019 Predictor_2	Bootstrap Forest		0,7182	0,8456	0,3228	4714
Pred Formula bin_ 019	Fit Least Squares		0,0066	1,5878	0,6018	4714

Model comparison

Response bin_ 019

Sorted Parameter Estimates

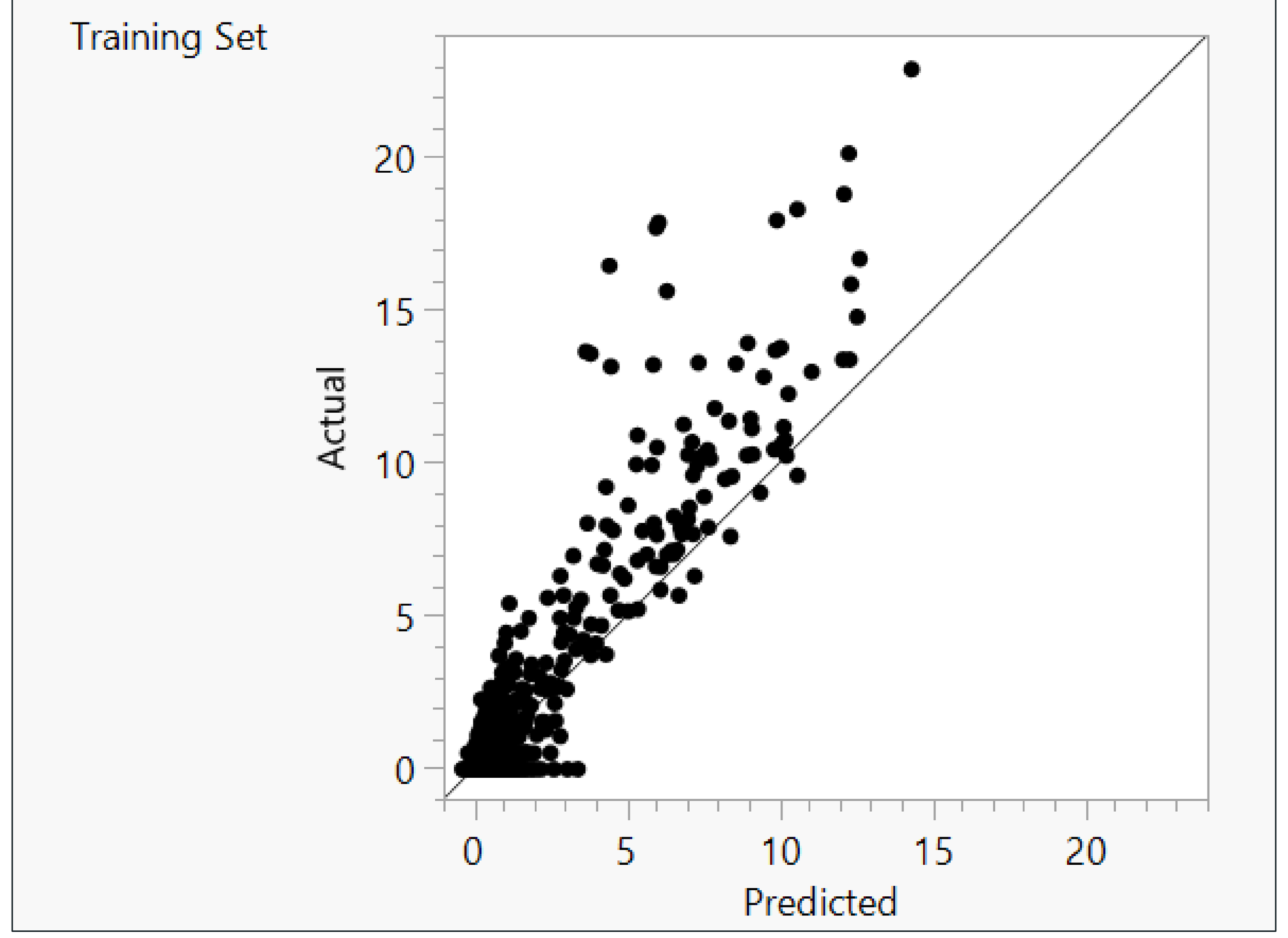
Term	Estimate	Std Error	t Ratio		Prob>|t|
test_ 273	1,775e-33	4,72e-34	3,76		0,0002*
test_ 001	-0,048096	0,018971	-2,54		0,0113*
test_ 424	-1,48e-12	7,17e-13	-2,06		0,0394*
test_ 143	0,0001862	0,000122	1,53		0,1271
test_ 265	-3,03e-34	4,99e-34	-0,61		0,5437
test_ 391	2,506e-33	4,42e-33	0,57		0,5708
test_ 457	-7,75e-33	1,59e-32	-0,49		0,6262
test_ 267	1,16e-33	2,94e-33	0,40		0,6927
test_ 422	4,414e-33	1,13e-32	0,39		0,6949
test_ 376	4,203e-33	1,13e-32	0,37		0,7088
test_ 268	-8,44e-34	1,59e-32	-0,05		0,9577

p-value of the most contributor tests provided by the Fit Least Square modeling: interest for this significancy result is relative

Multivariate analysis and modeling

- Four types of analysis are performed:
  - Boosted Tree
  - Partition
  - Boostrap Forest
  - Fit Least Squares
- Bin_019 is the response Y, the 11 tests selected each one from the 11 clusters are the X variables
- By comparing the 4 models, it appears that Boosted Tree is the best one.
- The result that the Device Engineers will use to understand the failing manufacturing step and to fix the yield loss, is the Column Contributions table

Actual by Predicted Plot



Actual by Predicted Plot in the Boosted Tree platform

Column Contributions

Term	Number of Splits	SS		Portion
test_ 143	625	12419,9352		0,3034
test_ 424	330	7599,19193		0,1856
test_ 001	367	6670,69103		0,1629
test_ 376	266	3950,70342		0,0965
test_ 267	186	3677,7064		0,0898
test_ 457	175	1508,56889		0,0368
test_ 391	192	1337,93963		0,0327
test_ 422	202	1210,48876		0,0296
test_ 265	113	1052,59832		0,0257
test_ 273	164	785,040562		0,0192
test_ 268	180	729,535957		0,0178

This Column Contributions table in the Boosted Tree platform is the main result: the Device Engineers will work on the manufacturing steps corresponding with the top 3 or 4 tests in this list

Root cause searching on a yield loss in automotive industry, by multivariate analysis and modeling (6/6)

EUROPE DISCOVERY SUMMIT

Conclusion / Next steps

From this analysis, the Device Engineers obtained some clues in their investigation for this yield loss issue: they were able to access a list of the CP tests seen as the most contributing to the yield loss observed on the bin of interest.
This result allowed them to save time in their research for the root cause of this yield issue.

Concern:

- For module 2 (Yield loss observed per UP test vs CP Test), need of the coordinate map between reticle and die for all mask sets from device engineers to be able to integrate CP and UP at reticle level

Next steps:

- Generate automated analysis flows for each analysis type
- Generate Design Templates for each analysis type
- Develop Interfaces
- Investigate possibilities to feed Final test data into Hadoop

Reference

Corinne Bergès[1], Shilu Zhang[2], Xiao Bai[2], Santosh Murali[3], Pete Smith[3], Jim Bird[2], Ivan Belen[1], Chris Smith[1],[1]NXP Quality Manufacturing/SPS, [2]NXP IT, [3]NXP ATMC Device Ops, 'CP-UP project _ Root cause searching on a yield loss issue, by multivariate analysis and machine learning', Toulouse Innovation Day (TID), 2023

Acknowledgements

The authors thank all the team members of this key project for NXP, for involving them in their activities and sharing their knowledge, thus allowing the development of machine learning in a new domain, i.e., root cause searching for yield loss or quality issues.

Thank you

Thank you for your attention

Questions / Contact

Contact: Corinne Bergès corinne.berges@nxp.com