10th National Convention on Statistics (NCS)
EDSA Shangri-La Hotel
October 1-2, 2007

Hybrid Regression:
A New Quality Control/ Process Control Tool

by

Jose H. Felizco Jr.

For additional information, please contact:

Author's name	:	Jose H. Felizco Jr.
Designation	:	P3-1 TPM/ 6-SIGMA Specialist
Affiliation	:	Amkor Technology Philippines
Address	:	Laguna Technopark, Sta.Rosa Laguna, Philippines
Tel. no.	:	(06349) 850-7000 loc 3853
E-mail	:	jfeli@amkor.com

Hybrid Regression:
A New Quality Control/Process Control Tool

by

Jose H. Felizco, Jr.

ABSTRACT

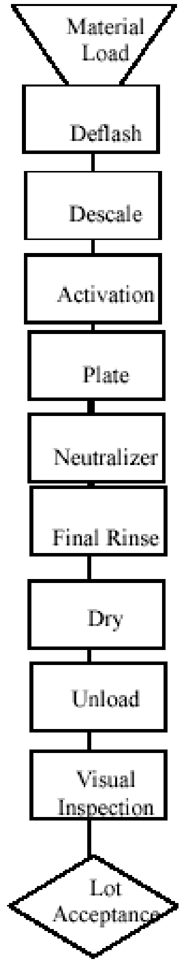
Prediction of motor temperature and conveyor misalignment in Central Lead Finish (CLF) area is essential for the sustenance of effective IR thermography equipment predictive maintenance and process control. It can be stated as a multivariate regression problem with real-time data. In this study, hybrid models of principal components regression (PCR) and partial least square (PLS) have been proposed. The basic idea of hybrid model is to combine the merits of PCR and PLS to develop more accurate prediction techniques. Both the components defined in PCR and the latent variables in PLS are involved in a hybrid model. Traditional PCR or PLS only extracts one kind of feature from the data maximum variance direction in PCR and maximum correlation direction with response variables in PLS. If these two kinds of features are included in a common regression model, the prediction performance is improved.

I. Introduction

Strip plating process is one of the most important and fundamental part of semiconductor assembly. An optimum plating process is indispensable because it minimizes ppm defective on the most vulnerable parts. Thus, monitoring and predicting condition of the plating process is important in order to make management decision on when to perform optimization. Prediction of ppm defective can be stated as a multivariate regression. The commonly used are principal component regression (PCR) and partial least square (PLS).

The objective of this study is to illustrate the merits of PCR and PLS, aiming at developing more accurate prediction techniques. Principal components (PC) defined in PCR as well as latent variables (LV) in PLS can be equated. Traditional PCR or PLS only extracts one kind of feature from the data i.e. maximum variance direction in PCR and maximum correlation direction with response variable in PLS. If these two kinds of features are included in a common regression model, the prediction performance is improved.

II.. Strip Plating System



Strip Plating Processes

Variables	Material	Condition
1. Load	N/A	Automatic=0, not auto=1
2. Deflash	1. Brand XYZ1 : 200-300 g/L	Temp: 40- 60 °C Current: 100-180 amps Measurement: ppm conformance
	2. Brand XYZ2 Sp. Gr.: 1.04 – 1.07	Temp: 40- 80 °C Current :: 100-180 amps Measurement ppm conformance
	3. Brand XYZ3 28 – 34 % by volume	Temp: 40- 60 °C Current: 100-180 amps Measurement: ppm conformance

	4. Brand XYZ4 : 250 – 500 g/L	Temp: 35- 50 ^{o}C Current: 100-180 amps Measurement: ppm conformance
3. Rinse	D.I. Water	Pressure: 5 psi min Flowrate: 30 L/h min Measurement: ppm conformance
4. High Pressure Rinse	D.I. Water	Pressure: 298 – 588 psi (20- 40 bars) Measurement: ppm conformance
5. Rinse	D.I. Water	Pressure: 5 psi min Flowrate: 30 L/h min Measurement: ppm conformance
6. Descale	1. Brand XYZ5: Salt: 10- 40 g/L Sulfuric Acid: 5- 15 % by vol Balance: D.I. Water	Temp: 20-30 ^{o}C Measurement: ppm conformance
	2. Brand XYZ6: Salt: 90 – 210 g/L Sulfuric Acid: 10- 40 ml/L Balance: DI Water	Temp: 30-40 ^{o}C Measurement: ppm conformance
	3. Brand XYZ7: Salt: 50 – 150 g/L Sulfuric Acid: 5 – 15 % by vol Balance: D.I. Water	Temp: 30-40 ^{o}C Measurement: ppm conformance

	4. Brand XYZ8: salt: 30 -50 g/l Sulfuric Acid: 10 - 30 ml /L balance DI water	Temp: 20-30 ^{o}C Measurement: ppm conformance
	5. Brand XYZ9: Salt : 30 – 50g/L	Temp : 30 – 40 ^{0}C Measurement: ppm conformance
	Autodosing Tank 6. Brand XYZ10: Salt: 200 - 270 g/L	Temp: 35- 50 ^{o}C
7. Rinse	D.I. Water	Pressure: 5 psi min Flowrate: 30 L/h min Measurement: ppm conformance
8. Activation	Acid: 10- 20 % by vol Balance: D.I. Water Acid : 8~12% by Vol for Matte tin	Temp: 35- 50 ^{o}C Current: 5-15 amps Measurement: ppm conformance
9. Plating-1	1. Plating Chemicals Tin : 35-55 g/L Acid : 30-40 % Primary Additive: 75-150 ml/L Secondary Additive: 6-10 ml/L Concentrate: 5-20 ml/L Balance: D.I. Water	Temp: 30-50 ^{o}C Current Density: 75 – 225 ASF Measurement: ppm conformance
10. Plating-2	Plating Chemicals: Tin: 230-315 g/L Primary Additive Sp. Gr: 1.0 –1.02 PH: < 2	Temp: 35- 50 ^{o}C Measurement: ppm conformance

11.Plating-3	Plating chemicals Tin : 35-55 g/l Acid : 30-40%	Temp : 30-50 ^{o}C Current : 75 – 225 ASF Measurement: ppm conformance
12. Plating-4	Plating Chemistry : 120 – 250 g/l Tin 11	Temp: 35 – 45 ^{o}C Current Density: 75 – 225 ASF Measurement: ppm conformance
13. Plating-5	Plating Chemicals-set A Acid : 8~12% by Vol Tin : 55 ~ 75 g/l Makeup : 6~10% by Vol Additive : 2~6% by Vol	Temperature : 35 - 45 ^{o}C Current : 75 – 225 ASF Measurement: ppm conformance
	Plating Chemicals-set B Tin : 65 – 85 g/L Acid HC: 255 – 285 ml/L Primary : 60 – 90 ml/L Secondary : 2 – 6 ml/L	Temp : 25 – 30 ^{0}C Current density : 75 - 225 ASF Measurement: ppm conformance
14. Rinse	D.I Water	Temp: 25 – 35 ^{o}C Measurement: ppm conformance
15. Neutralizer PST	Neutralizer chemicals: 1. Salt : 5-12 g/L pH: 8-12 Balance: DI Water	Temp: 35 – 55 ^{o}C Measurement: ppm conformance
	2. Potassium Phosphate Salt: 3-12 g/L Balance D.I. Water	Temp: 25 – 35 ^{o}C Measurement: ppm conformance
16. Neutralizer TSP	<u>Neutralizer Solution</u> Salt: 3-12 g/L Balance: D.I. Water	Temp: 35 – 55 ^{o}C Measurement: ppm conformance
17. Rinse	D.I. Water	Temp: 35- 50 ^{o}C

		Measurement: ppm conformance
18. Hot Rinse	D.I. Water	Temp: 55 – 75 ^{o}C Measurement: ppm conformance
19. Air Knives/ Blow Off	Compressed Dry Air (CDA)	Pressure: 0.4 – 0.5 bars Measurement: ppm conformance
20. Parts Dryer	Compressed Dry Air (CDA)	Temp: 110 – 130 ^{o}C Measurement: ppm conformance
21. Unload 22. Belt Stripper-1	N/A Stripper Solution Acid: 15- 35 % by vol Inhibitor (1L / 20L) Balance: D.I. Water	Automatic=0, not auto=1 Temp: 35- 50 ^{o}C Measurement: ppm conformance
23. Solder Stripper	Solder Stripper Nitric Acid: 25- 35% vol	Temp: 35- 50 ^{o}C Measurement: ppm conformance
24. Belt Stripper-2	3. Belt Stripper HN 980 B : 30 g/L	Temp: 50 – 60 ^{o}C Measurement: ppm conformance
25. Rinse	D.I. Water	Temp: 35- 50 ^{o}C Measurement: ppm conformance
26. Belt Dryer	Compressed Dry Air (CDA)	Temp: 100 – 120 ^{o}C Measurement: ppm conformance

III. Hybrid Regression vs. PCR/PLS

Prediction of ppm defective in plating area is essential for the sustenance of effective process control. It can be stated as a multivariate regression problem with real-time data. The basic idea of hybrid model is to combine the merits of PCR and PLS to develop more accurate prediction techniques. Both the components defined in PCR and the latent variables in PLS are involved in a hybrid model. Traditional PCR or PLS only extracts one kind of feature from the data maximum variance direction in PCR and maximum correlation direction with response variables in PLS. If these two kinds of features are included in a common regression model, the prediction performance is improved.

The results show that an optimal hybrid model can outperform PCR and PLS, especially when the number of predictor variables increases. It suggests that the proposed approach may be particularly useful for complex prediction tasks that need more predictor variables. Therefore, the key to hybrid model is that the linear transformed vector maybe either a principal component or latent variables.

IV. Data Analysis and Model Fitting

The data with several observations consists of plating ppm defective and various ppm conformance (process parameters' measurements) affecting conveyorized plating process. The data are taken from computer data log of plating area. The variables used are summarized in Table 1.

Table 1: Description of Variables

Variables	Descrption
Y	Ppm defective
X1	Load
X2	Electrolytic Deflash
X3	Rinse-1
X4	High Pressure Rinse
X5	Rinse-2
X6	Descale
X7	Rinse-3
X8	Activation
X9	Plating-1
X10	Plating-2
X11	Plating-3
X12	Plating-4
X13	Plating-5
X14	Rinse-4
X15	Neutralizer PST
X16	Neutralizer TSP
X17	Rinse-5
X18	Hot Rinse
X19	Air Knives/ Blow Off
X20	Parts Dryer

X21	Unload
X22	Belt Stripper-1
X23	Solder Stripper
X24	Belt Stripper-2
X25	Rinse-6
X26	Belt Dryer

Using the ordinary least square methods, coefficients and related statistics that have been calculated are presented in Table 2. Those calculations have been performed using computer software.

Table 2: Coefficients and Collinearity Statistics

Model	Unstandardize Coefficients		**Coefficient**	**t**	**Sig**	**Collinearity Stat**	
	B	**Std.Error**	**Beta**			**Tolerance**	**VIF**
Const.	-1714.5	528591.059		-0.0000	0.997		
X1							
X2	-1.043	0.821	-1.427	-1.27	0.21	0.004	267.175
X3	88.113	12543.582	0.005	0.007	0.994	0.011	94.617
X4	72.613	36.871	2.043	1.969	0.054	0.004	227.95
X5	-1239.6	664.815	-2.315	-1.865	0.068	0.003	326.386
X6	222.033	476.39	0.59	0.466	0.643	0.003	329.256
X7	328.163	3439.452	0.184	0.095	0.924	0.001	789.592
X8	-293.169	328.627	-1.056	-0.892	0.376	0.003	296.59
X9	88.738	44.571	3.429	1.991	0.052	0.002	628.148
X10	-2.826	3.24	-1.824	-0.872	0.387	0.001	925.401
X11	0.477	2.655	0.155	0.18	0.858	0.006	158.09
X12	-0.748	0.442	-1.012	-1.695	0.096	0.013	75.535
X13	7.026	1.884	0.824	3.729	0.000	0.097	10.341
X14	3.00E-02	0.013	0.405	2.302	0.025	0.153	6.538
X15	11.941	10.442	2.229	1.144	0.258	0.001	804.182
X16	-5.00E-03	0.005	-0.502	-1.002	0.321	0.019	53.24
X17	62.252	40.616	0.559	1.533	0.131	0.036	28.116
X18	22.224	13.366	4.584	1.633	0.302	0.001	1608.99

X19	6.558	10.431	0.606	0.629	0.532	0.005	196.703
X20	10.222	19.705	0.836	0.519	0.606	0.002	549.737
X21							
X22 X23	-8.00E-04 1.70E-02	0.004 0.008	-0.076 0.543	-0.233 2.142	0.817 0.037	0.044 0.073	22.659 13.611
X24	-8.00E-03	0.021	-0.075	-0.396	0.694	0.131	7.611
X25	-2.00E-02	0.024	-0.192	-0.972	0.336	0.121	8.256
X26	8387.467	6489.71	0.135	1.292	0.202	0.431	2.319

Looking at VIF column, only four of the independent variables (X14, X24, X25, X26) are significantly non-collinear, but the rest are highly correlated. Using ordinary least square, as ANOVA table has been performed to test whether the model, in which the regressors may be a linear combination of the predicted variable, is significant.

Table 3: Analysis of Variance for Strip Plating Conveyor Data

Model	Sum of Square	df	Mean Square	F	Sig
Regression	4.07E+1.3	26	1.57E+12	6.105	0.000
Residual	1.36E+13	53	2.56E+11		
Total	5.43E+13	79			

From Table 3, it is clear that the model is significant with a probability of 95%. Even though the OLS model fits the data well, multi-collinearity may severely prohibit quality prediction. RMSECV values for both PCR and PLS are calculated and plotted as a function of the number of latest variables.

Table 4 and Table 5 present the percent variance captured by the model. For the optimal number of latent variable in PCR, 100% of the variance is captured by the regressors. These latent variables (PCR) could explain 79.2% of the variation. This is equivalent to R-square in OLS. For the optimal number of latent variable in PLS, 100% of the variance is captured by the regressors. These latent variables (PLS) could explain 79.59% of the variation.

Table 4: Percent Variance Captured by Regression Model using PCR

	X block		Y block	
LV	This LV	Total	This LV	Total
1	86.05	86.1	53.69	53.7
2	5.15	94.5	2.06	55.8
3	6.25	97.6	19.63	75.4
4	2.09	99.6	3.6	79
5	0.28	99.8	0.11	79.1
6	0.15	100	0.05	79.1
7	0.04	100	0.04	79.2
8	0.00	100	0.47	79.7
9	0.00	100	1.98	81.6
10	0.00	100	1.51	83.1

Table 5: Percent Variance Captured by Regressors Model using PLS

	X block		Y block	
LV	This LV	Total	This LV	Total
1	86.05	86.05	55.63	55.63
2	5.15	91.19	19.54	75.17
3	6.25	97.44	3.6	78.77
4	2.09	99.53	0.28	79.05
5	0.28	99.81	0.12	79.16
6	0.15	99.96	0.07	79.24
7	0.04	100	0.35	79.59
8	0.00	100	3.23	82.82
9	0.00	100	0.44	83.26
10	0.00	100	3.97	87.23

It is evident from the comparison of the models that the fit and prediction are entirely different aspects of a model's performance. If prediction is the goal, a model that gives the minimum RMSECV value among the prediction models is selected.

In the comparisons of the models according to RMSE and RMSECV, plots were generated describing how well the model fits the data. These plots help to visualize the model performance for the prediction methods.

Table 6: RMSE and RMSECV Values for all Prediction Methods

	PCR	PCR'	PLS	PLS'
RMSE	506428	616250	591592	507285
RMSECV	1974920	1337700	1335270	1377270

V. Conclusions

As can be seen from table 6, the PCR model has the smallest RMSE value. The second smallest RMSE value belongs to the PLS' model. Under the conditions of no collinearity in the independent variables, this indicates that, the PCR and PLS' models fit the data better than both PCR' and PLS. For comparison of models intended for prediction it is inadequate to look just at model fit.

Future research includes creating nonlinear hybrid models (Hines 1999). A possible solution is based on Kernel Tricks (Scholkopf et.al 1998) which has been proven as efficient approach to deal with non-linear problem. Kernel PCR and kernel PLS have been proposed and achieved good prediction results (Rosipal 2001). Thus, it is expected that the kernel hybrid model of PCR and PLS could work well on some nonlinear cases.

Furthermore, model solution method has an exponential algorithm complexity. Although in many cases, the number of significant components is a small as 10 or fewer, it may be necessary in some instances to develop more efficient model selection method by constraining the search for candidate hybrid model.