J Intell Manuf (2012) 23:835–844
DOI 10.1007/s10845-010-0440-1

A non-linear quality improvement model using SVR for manufacturing TFT-LCDs

Der-Chiang Li · Wen-Chih Chen · Chiao-Wen Liu · Yao-San Lin

Received: 21 July 2009 / Accepted: 12 July 2010 / Published online: 5 August 2010

Abstract Thin Film Transistor—Liquid Crystal Displays (TFT-LCDs) are widely used in TVs, monitors, and PDAs. The key process of producing a TFT-LCD is using alignment to combine a Thin Film Transistor (TFT) panel with a Color Filter (CF) panel, which is called "celling". The defined cell vernier, which indicates the alignment error, is an important quality index in the manufacturing process. In the CF manufacturing process, the cell vernier is difficult to control because it depends on six TPEs (Total Pitch Errors), with each TPE highly dependent on the others. This paper aims to improve the cell vernier forecasting model with the six TPE attributes to enhance the production yield in the CF manufacturing process. Using the six dependent variables, this study found that the SVR (Support Vector Machine for Regression) model is the fittest for generating quality results that meet the designed specifications.

Keywords Dependent factor · SVR · Multi-regression · TFT-LCD · Color filter · Quality control

Introduction

Thin Film Transistor—Liquid Crystal Display (TFT-LCD) are widely used in TVs, notebooks, mobile phones, and digital cameras. A TFT-LCD mainly consists of a Thin Film Transistor (TFT) panel and a Color Filter (CF) panel. The process of producing a TFT-LCD is usually separated into three main steps, as shown in Fig. 1 (Chen et al. 2005): (i) the TFT and CF are produced using a process similar to that used for making semiconductors; (ii) the cells are made by combining a TFT panel with a CF panel, and then injecting Liquid Crystal (LC) into the chamber; (iii) the Modules are formed, which involves assembling polarizers and backlights.

In the process of making a cell, a TFT panel is precisely aligned with a CF panel to combine them. The cell vernier is defined as the distance (or error) between the TFT alignment marks and the CF alignment marks. If the value of the cell vernier is high, the panel will have insufficient light contrast. Insufficient light contrast decreases the price of the TFT-LCD. Hence, the cell vernier is a very important factor for high panel quality.

For a panel, manufacturers define the error between the designed and the actual pitch values as the total pitch error (TPE). There are six TPEs in a CF and six in a TFT. Considering the complex process of making a TFT, to reduce vernier values, engineers usually prefer to adjust the TPEs in a CF to match a TFT in the alignment process.

To analyze the cause-effect model of cell verniers, we obtained data from the SPC (statistic process control) database. Since testing cost is high, and only the most recent three months of production data were considered valid for the current manufacturing calibrations. Nineteen valid data points for making middle-size panels were collected; Table 1 shows the partial TPE raw data.

Generally, researchers use multi-regression as the tool for quality analysis when a robust multi-regression model can be constructed using linearly independent variables (Eicker & Ewald 2004; Chang & Cheng 2009; Chang et al. 2006). However, in cell vernier analysis, all of the input variables highly depend on each other. The present paper tested all the possible linear regression methods that on-line engineers can use for the cell vernier problem, such as building up a simple linear regression, using principal component analysis (PCA) to deal with the dependent variable problem, finding

D.-C. Li (✉) · W.-C. Chen · C.-W. Liu · Y.-S. Lin
Department of Industrial and Information Management, National Cheng Kung University, 1 University Road, Tainan 70101, Taiwan
e-mail: lidc@mail.ncku.edu.tw

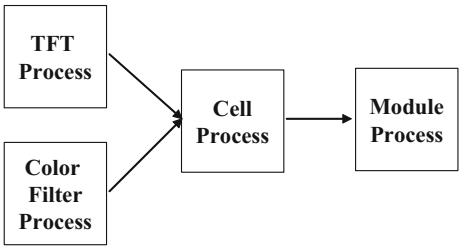

Fig. 1 TFT-LCD manufacturing process

Table 1 TPE raw data

TP_X1	TP_X4	TP_Y1	TP_Y6	TP_D1	TP_D2	y
0.73	1.00	1.12	1.47	2.39	0.68	10.20
−0.14	0.09	−0.68	−0.66	−0.13	−0.89	9.64
	.				.	
	.				.	
	.				.	
0.08	0.49	−0.11	−0.60	0.44	−0.58	9.93

the best subsets regression, and building up multi-regression. The support vector machine for regression (SVR) model was found to be the most appropriate for forecasting manufacturing performance.

Cross validation and Monte Carlo simulations were used to verify the forecasting models in the analyses. The experiment results show that the SVR is superior to the multi-regression models and that it provides confidence intervals for the manufacturing factors. This contribution will benefit engineers that need to determine TPE values to control the cell vernier values within the specifications.

The rest of this paper is organized as follows: Section "Problem description" describes the details problem. Traditional statistical analysis is given in Section "Traditional statistical analysis". Section "Support vector machine for regression (SVR) model" contains the support vector machine for regression model. The model verification is given in Section "Model verification". Finally, the discussion and conclusion are presented in Section "Model verification".

Problem description

For TFT-LCDs, the sale price depends on display quality, for which the cell vernier is the key factor. If the value of the cell vernier is out of specification, the product will be scraped.

The Cell vernier depends on the TPEs in TFT and CF processes. It is difficult to adjust the TPEs in TFT because the process of TFT is complex and the cost is high. It's thus assumed that the TPE values in TFT are fixed; the TPE values in CF are adjusted. In short, the cell vernier and TPE are the errors between the actual and standard coordinates. The main purpose of this study is to build a forecasting model which can predict cell verniers to control the relevant variables to retain high product quality.

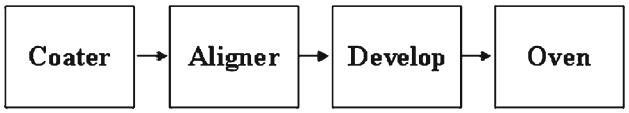

Fig. 3 Main procedure of fabricating BM/G/R/B/PS layers

Factors selection and data collection

We first introduce the manufacturing process for CF.

(1) Manufacturing process for CF

The CF process in this study is a flow-shop type procedure. The main processes are those for Black Matrix (BM), Green (G), Red (R), Black (B), Indium Tin Oxide (ITO), and Photo-Spacer (PS) layers as shown in Fig. 2. The function of the BM layer is to retain light, that of the G, R, and B layers is to display color, and that of the PS layer is to provide a gap between TFT and CF in the celling process. Except the ITO process, layers have similar procedures as shown in Fig. 3. In the coater stage, the resist is coated on the panel; in the aligner stage, the pattern is exposed; in the develop stage, wasted resist is removed; and in the oven stage, the solvent is vaporized.

In this case study, the CF alignment marks were produced in the BM aligner stage. In the G layer, R layer, B layer, PS layer, and celling processes glasses are aligned based on the alignment marks. The alignment stage in BM layer is the most important, and its alignment equipment is the most precise and accurate one.

(2) Factor selection

After consulting with the research and design (RD) engineers, it was found that the TFT process is more complex and stable than the CF process. The RD engineers thus assume that the TPE in TFT is fixed and adjust the TPE in CF to meet cell requirement.

In CF manufacturing, the cell verniers and TPEs are the errors between the actual and standard coordinates, so the values may be positive or negative. A panel in CF has twelve

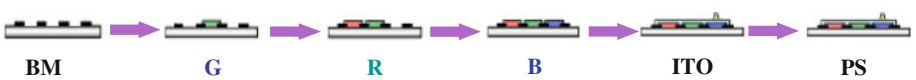

Fig. 2 Color Filter main processes

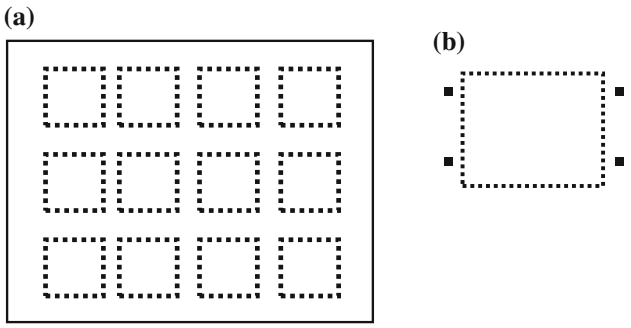

Fig. 4 **a** A panel with twelve panels. **b** A panel with four marks

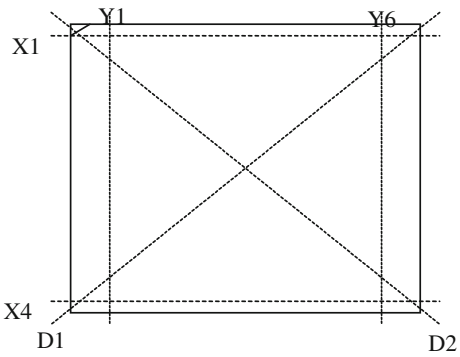

Fig. 5 Six TPE related coordinates in a panel

panels of nineteen inch, as shown in Fig. 4a. A panel has four marks as shown in Fig. 4b. A mark has two attributes, x and y axis coordinate values. A panel has ninety-six cell vernier values. The average of the absolute ninety-six cell vernier data is the output quality value. The input data is the six TPEs, $X1$, $X4$, $Y1$, $Y6$, $D1$, and $D2$, in a panel which are shown in Fig. 5. In the CF process, TPEs are collected and measured under a sampling of 1 out of 100.

Traditional statistical analysis

Regression and correlation

Minitab was used to find the correlation among the six variables. The results are shown in Table 2. The results reveal a high positive correlation between the factors. When a simple linear regression was tried on each TPE, the linear regression analysis failed (see Table 3), because the adjusted R-square of each equation is low and the P value of the Analysis of Variables (ANOVA) is not significant. Furthermore, the regression obtained using all factors in the model also failed because of the low adjusted R-square values and in-significant P value (see Table 4).

We next considered using stepwise regression to analyze the data, as described in the following section.

Table 2 Correlation of TPEs and cell verniers

	X1	X4	Y1	Y6	D1	D2	y
X1	1.00						
X4	0.50	1.00					
Y1	0.39	0.64	1.00				
Y6	0.81	0.53	0.71	1.00			
D1	0.70	0.71	0.83	0.88	1.00		
D2	0.73	0.74	0.79	0.84	0.75	1.00	
Y	−0.13	−0.03	0.19	−0.01	0.08	−0.02	1.00

Table 3 Linear regression of a single TPE and cell verniers

Simple regression analysis			
No.	Equation	R-Sq (adj) (%)	P value
1	Y = 1.403–0.1010X1	0.1	0.326
2	Y = 1.497–0.1626X4	13.2	0.070
3	Y = 1.343–0.0787Y1	2.5	0.243
4	Y = 1.376–0.1078Y6	4.1	0.200
5	Y = 1.439–0.1060D1	14.4	0.061
6	Y = 1.374–0.0692D2	0	0.363

Table 4 Regression with all TPEs

The regression equation $Y = 1.85 + 135X1 + 135X4 + 147Y1 + 147Y6 - 200D1 - 199D2$			P value 0.205
Predictor	Coef	P value	VIF
Constant	1.8468	0.000	
X1	135.10	0.113	773997.686
X4	134.93	0.114	946754.363
Y1	146.80	0.113	2091821.897
Y6	146.86	0.113	1334980.669
D1	−199.70	0.113	5174967.852
D2	−199.30	0.114	3088483.011

Stepwise regression and principal component analysis (PCA)

Since simple linear regression is not appropriate for analyzing the model, we used stepwise regression to find the factors which are statistically significant to the output.

To overcoming the high variance inflation factor (VIF) problem, Principal Component Analysis (PCA) was used to analyze the data (Hair et al. 1998), The principal component matrix are shown in Table 5, where the first three variables ($PC1$, $PC2$, and $PC3$) are chosen as the input variables with 95.7% explanation ability. The multiple-regression results with the three transformed variables $PC1$, $PC2$, and $PC3$ are shown in Table 6.

Table 5 Principal component matrix for TPEs

Variable	PC1	PC2	PC3	PC4	PC5	PC6
X1	0.361	−0.710	−0.134	−0.013	−0.539	−0.240
X4	0.393	0.342	−0.641	−0.487	0.102	−0.266
Y1	0.399	0.542	0.379	0.332	−0.381	−0.395
Y6	0.430	−0.285	0.347	0.013	0.717	−0.316
D1	0.439	0.060	0.393	−0.491	−0.148	0.621
D2	0.432	0.016	−0.391	0.641	0.136	0.480

Table 6 Multi-regression of PCA

The regression equation $Y = 1.41 - 0.050PC1 - 0.025PC2 - 0.005PC3$			*P* value 0.512
Predictor	Coef	*P* value	VIF
Constant	1.4105	0	
PC1	−0.0506	0.217	1.252
PC2	−0.0246	0.812	1.096
PC3	−0.0055	0.966	1.206

Table 6 reveals that the multiple-regression fails with an adjusted R-square value of almost zero. This is not good enough to model the data.

Our next option was to join the input factors step by step for the Multiple Regression model, as described in the following section.

Best subsets regression

We tried to estimate the multiple regressions using the nineteen data points, where each datum has six original factors. If we considered all the interactions of the factors, we would have more than one thousand factor combinations, which would prevent convergence. To avoid this problem, we first considered only the interaction of two-rank factors after consulting with the senior engineers in the factory. Thus, we use d the six original parameters and fifteen interaction parameters. With only nineteen data points, it is difficult to analyze data with twenty-one variables. Therefore, we selected some critical variables from the twenty-one variables using Best Subsets Regression in Minitab (Hocking 1976). We replaced the main original variables with $PC1$, $PC2$, and $PC3$ variables according to the results obtained in Secttion "Stepwise regression and principal component analysis (PCA)". We also replaced $X1 \times Y6$, $X4 \times Y1$, and $X4 \times Y6$ interaction variables with $D1 \times D2$ for the analysis (see Table 7). Finally, only fifteen variables ($PC1$, $PC2$, $PC3$, $X1 \times D1$, $X1 \times D2$, $X4 \times D1$, $X4 \times D2$, $Y1 \times D1$, $Y1 \times D2$, $Y6 \times D1$, $Y6 \times D2$, $X1 \times X4$,$Y1 \times Y6$, $D1 \times D2$, $X1 \times Y1$) were used to analyze the data in the Best Subsets Regression. The results are shown in Table 8. When choosing the row cut of vars. = 13 and CP = 12.1 as the representative model, the 13 critical variables are $PC1$, $PC2$, $PC3$, , $X1 \times D2$, $X4 \times D2$, $Y1 \times D1$, $Y1 \times D2$, , $X1 \times X4$, $Y1 \times Y6$, $D1 \times D2$, and $X1 \times Y1$.

Table 7 Multi-regression: D1*D2 versus X1*X6, X4*Y1, and X4*Y6

The regression equation $D1 * D2 = 0.248 + 0.499X1 * Y6 + 0.420X4 * Y1 + 0.674X4 * Y6$			*P* value <0.01
Predictor	Coef	*P* value	VIF
Constant	0.2480	0.048	
X1*Y6	0.4994	0.003	3.234
X4*Y1	0.4198	0.004	3.345
X4*Y6	0.6738	0.005	6.837

Multiple regressions

After the critical variables in the model were estimated in the Best Subset Regression, we tried to reduce the model using the Multiple Regression method.

First, we input all 13 critical variables into the multiple regressions. Only two significant variables, $PC2$ ($p = 0.045$) and $X1 \times Y1$ ($p = 0.028$), were found (see Table 9). We thus reset the model significance level from $\alpha = 0.05$ to $\alpha = 0.1$ and chose $PC2$, $X1 \times D1$, $X1 \times D2$, $Y1 \times D1$, and $X1 \times Y1$ to execute the multi-regression (see Table 10). Because the p-value of variables was less than 0.1, we further reduced the model variables to $PC2$, $X1 \times D1$, and $X1 \times Y1$ (see Table 11). In Table 11, non of the variables are significant, so we further reduced the model variables to $PC2$ and $X1 \times D1$. Unfortunately, we again found no significant variable in the model (see Table 12).

In Section "Stepwise regression and principal component analysis (PCA)", we tried to overcome the VIF problem by using PCA but the results show no significant variable in the multi-regression. To keep the characteristics of the original data, we next tried to use the original main variables to replace the PCA variables, $PC1$, $PC2$, and $PC3$, to develop the multi-regression. Finally we established the model with the original main variables and the interaction variables found in Section "Best subsets regression". The Multiple Regression is shown in Table 13.

However, some of the variables in the multiple regressions still have high p values, which indicate low significance. Therefore, we reduced the model in Multiple Regression by removing the variables with high p values to obtain the final model using variables $X1$, $X4$, $Y1$, $Y6$, $D1$, $D2$, $Y1 \times Y6$, $D1 \times D2$, $X1 \times D1$, and $X4 \times D2$ (see Table 14). That is:

Table 8 Best subset regression of the selected variables

Best Subsets Regression: Y versus Pc1, PC2,…

Response is Y

Mallows					PC	PC	PC	X1*D	X1*D	X4*D	X4*D	Y1*D	Y1*D	Y6*D	Y6*D	X1*X	Y1*Y	D1*D	X1*Y
Vars	R-Sq	R-Sq (adj)	Cp	S	1	2	3	1	2	1	2	1	2	1	2	4	6	2	1
1	15.1	10.1	−2.4	0.27566						X									
1	14.5	9.5	−2.3	0.27668									X						
2	24.1	14.6	−1.7	0.26866														X	X
2	22.7	13.0	−1.5	0.27120									X						X
3	35.4	22.4	−1.4	0.25611	X														X
3	34.9	21.9	−1.3	0.25704												X		X	X
4	43.6	27.5	−0.6	0.24757		X			X		X							X	X
4	43.2	27.0	−0.6	0.24849		X						X			X				X
5	50.9	32.1	0.3	0.23965			X							X			X	X	X
5	48.3	28.4	0.7	0.24607		X	X						X						X
6	52.9	29.3	2.0	0.24451			X	X			X			X	X		X	X	X
6	52.5	28.8	2.0	0.24536			X						X	X			X	X	X
7	63.9	40.9	2.4	0.22361	X	X		X	X			X				X			X
7	61.7	37.3	2.7	0.23029	X	X		X	X	X		X							X
8	70.7	47.3	3.3	0.21104	X	X		X	X	X		X				X		X	X
8	65.4	37.8	4.1	0.22940	X	X		X	X			X				X			X
9	72.9	45.8	5.0	0.21410	X	X		X	X		X	X				X		X	X
9	72.0	44.0	5.2	0.21762	X	X		X	X			X				X		X	X
10	75.5	44.9	6.6	0.21593	X	X		X	X		X	X				X	X	X	X
10	74.0	41.5	6.9	0.22246	X	X		X	X		X	X	X			X		X	X
11	77.9	43.1	8.3	0.21936	X	X		X	X		X	X	X			X	X	X	X
11	77.7	42.6	8.3	0.22033	X	X		X	X		X	X			X	X	X	X	X
12	78.7	36.0	10.2	0.23257	X	X		X	X		X	X	X		X	X	X	X	X
12	78.7	36.0	10.2	0.23266	X	X	X	X	X		X	X			X	X	X	X	X
13	79.6	26.4	12.0	0.24948	X	X	X	X	X		X	X		X	X	X	X	X	X
13	79.2	25.0	12.1	0.25181	X	X	X	X	X		X	X	X		X	X	X	X	X
14	79.6	8.3	14.0	0.27847	X	X	X	X	X		X	X	X	X	X	X	X	X	X
14	79.6	8.1	14.0	0.27868	X	X	X	X	X	X	X	X		X	X	X	X	X	X
15	79.8	0.0	16.0	0.32030	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

$$Y = 2.09 + 255X1 + 257X4 + 278Y1 + 279Y6 - 379D1 - 378D2 + 0.639Y1 \times Y6 - 1.30D1 \times D2 + 1.17X1 \times D1 + 0.500X4 \times D2 \quad (1)$$

Regression model (1) was accepted since it fits the regression hypothesis. However, the VIF was still high, so we tried to use other methods to analyze the data, such as support vector regression.

Support vector machine for regression (SVR) model

SVR algorithm

Support vector machine (SVM) is a promising pattern recognition technique proposed 1995 by Vapnik et al. Unlike traditional methods which minimize the training error, SVM aims at minimizing an upper bound of the generalization error by maximizing the margin between the separating hyperplane and the data (Shawkat et al. 2006; Amari & Wu 1999). In

Table 9 Multi-regression with PCA and interaction variables

The regression equation			*P* value
$Y = 1.92 + 0.652PC1 - 1.880PC2 - 0.130PC3$ $-1.260X1 * D1 - 2.180X1 * D2 - 0.552X4 * D2$ $-0.918Y1 * D1 - 0.125Y1 * D2 + 0.383Y6 * D2$ $+0.748X1 * X4 - 0.643Y1 * Y6 + 1.360D1 * D2$ $+2.310X1 * Y1$			0.356
Predictor	Coef	*P* value	VIF
Constant	1.9240	0.001	
PC1	0.6523	0.116	132.055
PC2	−1.8829	0.045	73.859
PC3	−0.1300	0.745	14.558
X1*D1	−1.2577	0.068	216.585
X1*D2	−2.1811	0.073	373.588
X4*D2	−0.5521	0.272	121.596
Y1*D1	−0.9184	0.071	85.313
Y1*D2	−0.1248	0.741	26.620
Y6*D2	0.3828	0.601	139.616
X1*X4	0.7484	0.116	50.510
Y1*Y6	−0.6432	0.344	52.684
D1*D2	1.3621	0.198	640.828
X1*Y1	2.3082	0.028	136.640

Table 10 First reduction model of multi-regression

The regression equation			*P* value
$Y = 1.56 - 0.260PC2 - 0.100X1 * D1 - 0.142X1 * D2$ $-0.157Y1 * D1 + 0.479X1 * Y1$			0.193
Predictor	Coef	*P* value	VIF
Constant	1.5644	0.000	
PC2	−0.2596	0.085	2.574
X1*D1	−0.1004	0.302	5.809
X1*D2	−0.1419	0.260	5.293
Y1*D1	−0.1573	0.082	3.334
X1*Y1	0.4791	0.030	8.347

Table 11 Second reduction model of multi-regression

The regression equation			*P* value
$Y = 1.41 - 0.251PC2 - 0.167X1 * D1 + 0.247X1 * Y1$			0.250
Predictor	Coef	*P* value	VIF
Constant	1.4126	0.000	
PC2	−0.2512	0.096	2.396
X1*D1	−0.1667	0.066	4.223
X1*Y1	0.2465	0.140	4.854

its original form, SVM learning leads to a quadratic program which is a convex constrained optimization problem and thus has a unique solution. For a classification problem, given a training set of N samples: $(\mathbf{x}_1, t_1)$, $(\mathbf{x}_2, t_2)$, $\ldots$, $(\mathbf{x}_N, t_N)$,

Table 12 Third reduction model of multi-regression

The regression equation			*P* value
$Y = 1.39 - 0.085PC2 - 0.053X1 * D1$			0.397
Predictor	Coef	*P* value	VIF
Constant	1.3869	0.000	
PC2	−0.0845	0.396	1.029
X1*D1	−0.0529	0.238	1.029

Table 13 Multi-regression with main and interaction variables

The regression equation			*P* value
$Y = 2.06 + 266X1 + 268X4 + 290Y1 + 290Y6$ $-395D1 - 394D2 + 1.07X1 * D1$ $-0.06X1 * D2 + 0.425X4 * D2$ $+0.128Y1 * D1 - 0.087Y1 * D2$ $+0.199Y6 * D2 + 0.043X1 * X4$ $+0.423Y1 * Y6 - 1.30D1 * D2$ $+0.072X1 * Y1$			0.032
Predictor	Coef	*P* value	VIF
Constant	2.0643	0.001	
X1	266.16	0.013	2574267.318
X4	267.55	0.013	3194616.607
Y1	290.15	0.013	7070252.352
Y6	290.38	0.013	4461061.358
D1	−395.32	0.013	1.7447E+07
D2	−393.85	0.013	1.0391E+07
X1*D1	1.0676	0.058	1102.847
X1*D2	−0.0603	0.864	808.131
X4*D2	0.4249	0.122	336.784
Y1*D1	0.1283	0.486	250.603
Y1*D2	−0.0869	0.512	52.088
Y6*D2	0.1986	0.348	164.254
X1*X4	0.0430	0.782	124.714
Y1*Y6	0.4234	0.150	98.862
D1*D2	−1.2995	0.065	1897.319
X1*Y1	0.0719	0.826	405.551

where $\mathbf{x}_i \in \mathbb{R}^M$ is the input vector corresponding to the i^{th} sample labeled by $t_i \in \{-1, +1\}$ depending on its class, the SVM problem can be formulated as a quadratic programming optimization model to find the weight parameter $\mathbf{w}$ and the bias parameter b that maximize the margin while ensuring that the training samples are well classified.

$$\begin{aligned} \min \quad & \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{N}\xi_i \\ \text{s.t } & t_i\left(\mathbf{w}^{\mathrm{T}}\cdot\varphi(\mathbf{x}_i) + b\right) \geq 1 - \xi_i \\ & \xi_i \geq 0 \qquad i = 1, \ldots, N \end{aligned}$$

Table 14 Final multi-regression of main and interaction variables

The regression equation			*P* value
Y = 2.09 + 255X1 + 257X4 + 278Y1 + 279Y6 −379D1 − 378D2 + 0.639Y1 ∗ Y6 −1.30D1 ∗ D2 + 1.17X1 ∗ D1 + 0.5X4 ∗ D2			<0.01
Predictor	Coef	*P* value	VIF
Constant	2.0943	0.000	
X1	255.11	0.000	936485.722
X4	256.63	0.000	1150649.442
Y1	278.31	0.000	2538258.443
Y6	278.65	0.000	1619964.71
D1	−379.26	0.000	6286854.815
D2	−377.79	0.000	3745258.011
Y1*Y6	0.6388	0.000	22.076
D1*D2	−1.3015	0.000	177.704
X1*D1	1.1663	0.000	82.772
X4*D2	0.4996	0.001	78.598

Here, parameter C is used to tune the acceptable amount of errors. $\xi_i = |t_i - y(\mathbf{x}_i)|,\ i = 1, \ldots, n$ are slack variables and $y(\mathbf{x}_i) = \mathbf{w}^T \cdot \varphi(\mathbf{x}_i) + b$,

Cortes & Vapnik (1995) extended the SVM to regression models for treating a regression problem as a single classification case. Given a training set of N samples: $(\mathbf{x}_1, t_1),\ (\mathbf{x}_2, t_2), \ldots,\ (\mathbf{x}_N, t_N)$, where $\mathbf{x}_i \in \mathbb{R}^M$ is the input vector and the corresponding value$t_i \in \mathbb{R}$ is the target value of $\mathbf{x}_i$, and an ε-insensitive error function (the dotted line in Fig. 7) is proposed to adjust the amount of error.

$$E_\varepsilon(y(\mathbf{x}) - t) = \begin{cases} |y(\mathbf{x}) - t| - \varepsilon, & \text{otherwise} \\ 0, & \text{if } |y(\mathbf{x}) - t| < \varepsilon \end{cases}$$

The objective function (also called the error function) can be minimized using;

$$C \sum_{i=1}^{N} E_\varepsilon(y(\mathbf{x}_i) - t_i) + \frac{1}{2}||\mathbf{w}||^2$$

We can re-express the optimization problem by introducing slack variables to the model. For each data point $\mathbf{x}_i$, we use two slack variables, $\xi_i \geq 0$ and $\hat{\xi}_i \geq 0$, to outline the points that are out of the interval $[y - \varepsilon,\ y + \varepsilon]$, where $\xi_i > 0$ corresponds to a point for which $t_i > y(\mathbf{x}_i) + \varepsilon$, and $\hat{\xi}_i > 0$corresponds to a point for which $t_i < y(\mathbf{x}_i) + \varepsilon$, as illustrated in Fig. 6.

The condition for a target point to lie inside the ε-tube is that $y(\mathbf{x}_i) - \varepsilon \leq t_i \leq y(\mathbf{x}_i) + \varepsilon$. The purpose of introducing the slack variables is to allow points to lie outside the tube provided that the slack variables are non-zero, and the corresponding conditions are:

$$\begin{aligned} t_i &\leq y(\mathbf{x}_i) + \varepsilon + \xi_i \\ t_i &\geq y(\mathbf{x}_i) - \varepsilon - \hat{\xi}_i\ , \quad i = 1, \ldots, N \end{aligned}$$

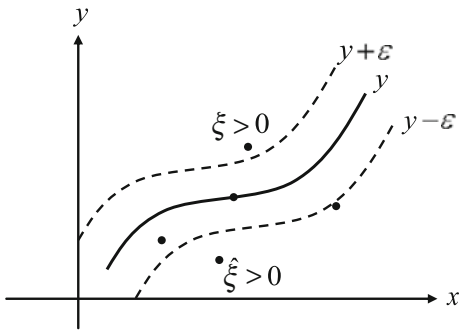

Fig. 6 Illustration of SVM regressions

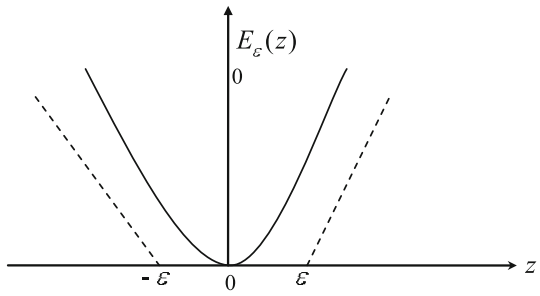

Fig. 7 Comparison of ε-insensitive error function (*dotted line*) and quadratic error function (*solid line*)

Hence, the error function for support vector regression can be written as :

$$C \sum_{i=1}^{N} (\xi_i + \hat{\xi}_i) + \frac{1}{2}||\mathbf{w}||^2$$

Using the Lagrange multipliers $\alpha_i \geq 0,\ i = 1, \ldots, N$ to solve the quadratic programming problem, we can find the dual problem as:

$$\begin{aligned} \max \quad & -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j) k(\mathbf{x}_i, \mathbf{x}_j) \\ & -\varepsilon \sum_{i=1}^{N} (\alpha_i + \hat{\alpha}_i) + \sum_{j=1}^{N} (\alpha_j - \hat{\alpha}_j) t_j \\ \text{s.t} \quad & 0 \leq \alpha_i \leq C, \\ & 0 \leq \hat{\alpha}_i \leq C, \quad i = 1, \ldots, N \end{aligned}$$

where $k(\cdot\ ,\ \cdot)$ is called a positive semidefinite kernel (or Mercer kernel) which satisfies the symmetric property (i.e. $k(\mathbf{x}_i, \mathbf{x}_j) = k(\mathbf{x}_j, \mathbf{x}_i)$) and the following equation:

$$\sum_{i=1}^{p} \sum_{j=1}^{p} b_i b_j k(\mathbf{x}_i, \mathbf{x}_j) \geq 0 \quad \forall\ p \geq 2,$$

$$\text{where } b_q \in \mathbb{R}\ \forall q = 1, \ldots, p.$$

Each Mercer kernel can be expressed as $k(\mathbf{x},\ \mathbf{x}') = \langle \phi(\mathbf{x}),\ \phi(\mathbf{x}') \rangle$, where $\phi : X \rightarrow F$ is a mapping of feature selection, and $\langle \cdot\ ,\ \cdot \rangle$ is the inner product. Commonly used

kernel functions include polynomial kernels, radial-basis function kernels, and two-layer perceptron kernels (Haykin 1999), respectively expressed as:

$$k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^t\mathbf{y} + 1)^d$$
$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{||\mathbf{x}-\mathbf{y}||^2}{2\sigma^2}\right)$$
$$k(\mathbf{x}, \mathbf{y}) = \tanh(\beta_0\mathbf{x}^t\mathbf{y} + \beta_1)$$

where d, σ, β_0, and β_1 are specified as a priori by the user.

Many studies have discussed the performance of SVR for various kernels (Sánchez 2003; Shawkat et al. 2006). Shawkat et al. indicated that the radial-basis function kernels have superior performance thus use this type of kernel as the analysis tool in this paper.

SVR model building for cell vernier

SVR is a kernel method for regression based on the principle of structural risk minimization (Smola & Scholkoph 2004). We used SVR to make a nonlinear prediction. In this research, we used this regression in WEKA software to analyze the nineteen data points with the Radial Basis Function (RBF as the kernel function (Witten & Frank 2005). We obtained the model function:

Support Vector Expansion:

$$\begin{aligned} Y = {} & (-1) \times K[X(0), X] + (1) \times K[X(1), X] + (-1) \\ & \times K[X(2), X] + (1) \times K[X(3), X] \\ & + (1) \times K[X(4), X] + (1) \times K[X(5), X] + (-1) \\ & \times K[X(7), X] + (-1) \times K[X(8), X] \\ & + (1) \times K[X(9), X] + (-1) \times K[X(10), X] + (1) \\ & \times K[X(11), X] + (-1) \times K[X(12), X] \\ & + (1) \times K[X(13), X] + (1) \times K[X(14), X] + (-1) \\ & \times K[X(15), X] + (-1) \times K[X(16), X] \\ & + (1) \times K[X(17), X] + (-1) \times K[X(18), X] + 1.3176 \end{aligned} \quad (2)$$

Kernel:

RBF kernel:

$$K(x, y) = e^{-(0.01 \times <x-y, x-y>^2)} \quad (3)$$

The mean absolute error is 0.1828 which is considered good enough to apply SVR to predict cell verniers.

Model verification

Since the TFT - LCD market is growing rapidly, it is difficult to spare manufacturing capacity to conduct an experiment. We can thus only verify and compare the two models, Multi-Regression and SVR, with simulated data.

Table 15 Cross-validation of multi-regression

	Y act	Y pied	Diff	Diff (%)	R-sq (%)	*P*-value (%)
All					98.90	0.00
19–1	1.1087	1.2398	−0.131	−11.83	99.00	0.01
19–2	1.5990	1.6840	−0.085	−5.31	98.93	0.01
19–3	1.1971	1.2691	−0.072	−6.02	98.99	0.01
19–4	1.3663	1.4849	−0.119	−8.68	99.08	0.00
19–5	1.2692	1.3071	−0.038	−2.99	98.92	0.01
19–6	1.8788	1.8249	0.054	2.87	98.67	0.01
19–7	1.1183	1.0469	0.071	6.38	98.98	0.01
19–8	1.0731	1.1986	−0.125	−11.69	99.19	0.00
19–9	1.0913	1.1968	−0.105	−9.66	99.03	0.00
19–10	1.5067	1.4063	0.100	6.67	98.92	0.01
19–11	0.9798	1.1367	−0.157	−16.01	98.93	0.01
19–12	1.2962	1.2800	0.016	1.25	98.91	0.01
19–13	0.9471	0.8768	0.070	7.43	98.87	0.01
19–14	1.9577	1.9118	0.046	2.34	98.51	0.02
19–15	1.4135	1.3269	0.087	6.13	99.01	0.01
19–16	1.2865	1.6111	−0.325	−25.22	99.26	0.00
19–17	1.1923	1.0280	0.164	13.78	99.42	0.00
19–18	1.6904	1.6682	0.022	1.31	98.79	0.01
19–19	1.0894	0.9897	0.100	9.15	99.02	0.00

(1) Interactive proof of the Multi-Regression model

Since the VIF of the regression was high (as described in Section "Stepwise regression and principal component analysis (PCA)"), we used cross validation to verify the model (Tan et al. 2006). We used 18 data points to fit the regression, and used the last data point to test the regression. EXCEL VBA software was used in the simulation. The results are shown in Table 15. In the table, "Diff" is the difference between the actual y and the predicted y; "Diff%" is diff divided by the actual y; "R-sq" is the R-square value of the multi-regression; and "*P*-value" means the *p*-value of ANOVA in the model.

The cell vernier specification is normally between 0 and 2. If the cell vernier is more than 2, the quality of the cell decreases, which lowers its sale price. From the data in Table 15, the regression model is considered reliable.

(2) SVR and Multi-Regression with Monte Carlo simulations

We simulated the CF TPE data sets under a normal distribution with six distinct sample means and variances (see Table 16). We simulated 10,000 sets of data to predict the cell verniers. The results show that only twenty seven data sets are within the specification for the multi-regression model. The rate of meeting the specification of SVR is 99.68%. Therefore, SVR is considered superior to multi-regression.

Table 16 Simulation data set with normal (mean, standard)

	X1	X4	Y1	Y6	D1	D2
Mean	0.83	1.10	0.30	0.53	1.13	0.79
Standard	0.69	0.76	1.04	0.83	1.20	0.93

(3) Using the research results in factories

To improve TFT-LCD quality, we used the simulated SVR data to set the parameters in CF. We selected all the cell vernier data 0 and 0.15. The corresponding parameters for a confidence interval of 95% are:

$X1$ in CF TPE is set (0.12349, 0.32249).
$X4$ in CF TPE is set (1.8398, 2.0471).
$Y1$ in CF TPE is set (0.6368, 1.03259).
$Y6$ in CF TPE is set (0.0966, 0.33695).
$D1$ in CF TPE is set (2.1670, 2.4275).
$D2$ in CF TPE is set (−0.1367, 0.1786).

We believe that if the CF TPE data conform to the above parameter confidence intervals in CF, the cell vernier will be within the specifications. In practice, it is difficult to modulate all six CF TPEs in the CF to meet the specified value. Engineers usually gradually adjust the CF TPE data in the CF one by one during CF production. Engineers in CF will adjust three or four TPEs of the six TPEs at a time to meet the confidence intervals in mass production, and leave other factors for the next adjustment.

Discussion and conclusion

This research analyzes the data step by step using correlation analysis, PCA, multi-regression, and SVR. We find that SVR and multi-regression are both appropriate models, with SVR more suitable than multi-regression and finally use the Monte Carlo simulation technique to fine the confidence interval for all the attributes to the on-line engineers. The proposed model helps online engineers, in a short period of time, to set the best parameters in cell vernier and improve the production yield. In practice (really applied in the factory), the proposed model finds the optimal TPEs in CF to improve the cell vernier significantly. However, for business consideration, the details can not be fully shown to the public at this moment. For this, an acknowledgement to the manufacturer is added to the paper at the end.

This research mainly aims to solve a problem with small data sets and dependent variables. Different from the past literatures to deal with small data set problem, we do not generate virtual samples or synthetic attributes to increase data information to achieve the objective of yield improvement. We develop a relatively engineer-applicable method instead of those research-oriented complex algorithms. The contributions of this study are three: first one is to test all the possible linear regression methods that on-line engineers can use for the cell vernier problem and finally show the SVR model is the most appropriate one for forecasting manufacturing performance; second is using Monte Carlo simulation technique to fine the confidence interval for all the attributes for engineers in a short period of time; third is the proposed model can be implement to all the TFT-LCD companies in their manufacturing systems.

In the future study, solving the interaction effect of variables in regression for dependent data will be one of the research directions. The virtual sample generation methods such as Niyogi et al. (1998), Li et al. (2003, 2006), Huang & Moraga (2004), and Li & Lin (2006) used prior knowledge, probability density functions, or information diffusion theories combined with neural networks to generate virtual data can be also considered in the future research direction.

Acknowledgments This research thanks for the HannStar Display Corporation Company's support and the data set providing.

References

Amari, S., & Wu, S. (1999). Improving support vector machine classifiers by modifying kernel functions. *Neural Networks, 12*, 783–789.
Chang, K.-H., Cheng, C.-H. (2009). Evaluating the risk of failure using the fuzzy OWA and DEMATEL method. *Journal of Intelligent Manufacturing*, Published online.
Chang, P.-C., Chen, L.-Y., & Liu, C. H. (2006). An evolutionary regulation algorithm for the twin laser measuring system. *Journal of Intelligent Manufacturing, 17*(5), 545–556.
Chen, S.-H., Chang, C.-K., & Kuo, C.-L. (2005). Optimal scheduling in color filter manufacturing. *Journal of the Chinese Institute of Industrial Engineers, 22*(4), 301–308.
Cortes, C., & Vapnik, V. N. (1995). Support vector networks. *Machine Learning, 20*, 273–297.
Eicker, F., & Ewald, G. (2004). Linearly independent set families. *Linear Algebra and Its Applications, 388*(1), 173–191.
Hair, J.F. Jr., Ralph, E.A., Ronald, L.T., & William, C.B. (1998). *Multivariable data analysis 5th*. Upper Saddle River, NJ: Prentice Hall.
Haykin, S. (1999). *Neural Networks—a comprehensive foundation*. Upper Saddle River, NJ: Prentice Hall.
Hocking, R.R. (1976). A biometrics invited paper: The analysis and selection of variables in linear regression. *Biometrics, 32*, 1–49.
Huang, C.F., & Moraga, C. (2004). A diffusion-neural-network for learning from small samples. *International Journal of Approximate Reasoning, 35*, 137–161.
Li, D.C., & Lin, Y.S. (2006). Using virtual sample generation to build up management knowledge in the early manufacturing stages. *European Journal of Operational Research, 175*, 413–434.
Li, D.C., Wu, C.S., & Chang, F.M. (2006). Using data continualization and expansion to improve small data set learning accuracy for early flexible manufacturing system (FMS) scheduling. *International Journal of Production Research, 44*(21), 4491–4509.
Li, D.C., Chen, L.-S., & Lin, Y.-S. (2003). Using functional virtual population as assistance to learn scheduling knowledge in

dynamic manufacturing environments. *International Journal of Production Research, 41*(17), 4011–4021.

Niyogi, P., Girosi, F., & Tomaso, P. (1998). Incorporating prior information in machine learning by creating virtual examples. In: *Proceedings of the IEEE* (pp. 275–298).

Sánchez, A.V.D. (2003). Advanced support vector machines and kernel methods. *Neurocomputing, 55*, 5–20.

Shawkat, A., Kate, A., & Smith, M. (2006). A meta-learning approach to automatic kernel selection for support vector machines. *Neurocomputing, 70*, 173–186.

Smola, A.J., & Scholkoph, B. (2004). A tutorial on support vector regression. *Statistic Computing, 14*(3), 199–222.

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. New York: Pearson Press.

Witten, I.H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques 2nd*. San Francisco: Morgan Kaufmann.