



# **A Study into the Deviation in Actual Printer Usage from the Expected and Deriving the Ideal Coverage for a Device (CSI ZG628T: Dissertation)**

by  
**Stephen Mathew**

Dissertation work carried out at  
Xerox Technology Services India LLP, Kochi, Kerala  
November 2019

Submitted in partial fulfillment of M.Tech. Computing Systems and Infrastructure degree programme

# Scope of Work

---



- Study the actual printer usage over the past year for Xerox's customer's devices and compare this with the calculated usage which was expected.
- Based on the comparison of actual and expected, come up with a data model to track the ideal coverage
- Implement this data model in a programming language to find the Ideal Coverage specific for a device based on its usage comparison

# Approach

---



## ***Formulae used***

- Calculate the Calculated Used and Actual Usage for toners
- Comparing Calculated Used and Actual Usage
- $f(x) = x \cdot b_1 + b_0$
- $x = \text{Usage Ratio based on Counters since last Swapout}$
- $y = 100 - PMS$
- $\text{Ideal Coverage (for a consumable)} = b_1 * \text{Rated Coverage}$

## ***Data collection and preprocessing***

- Data collection using SQL – was challenging
- One customer's data used for analysis
- Data cleaning in Excel:
  - cases where printers had meter reads and supply levels for one or more consumables, but either meter reads or supply levels for a specific date missing
  - Data from multiple reference points were considered
- Pandas DataFrames were used for some data preprocessing

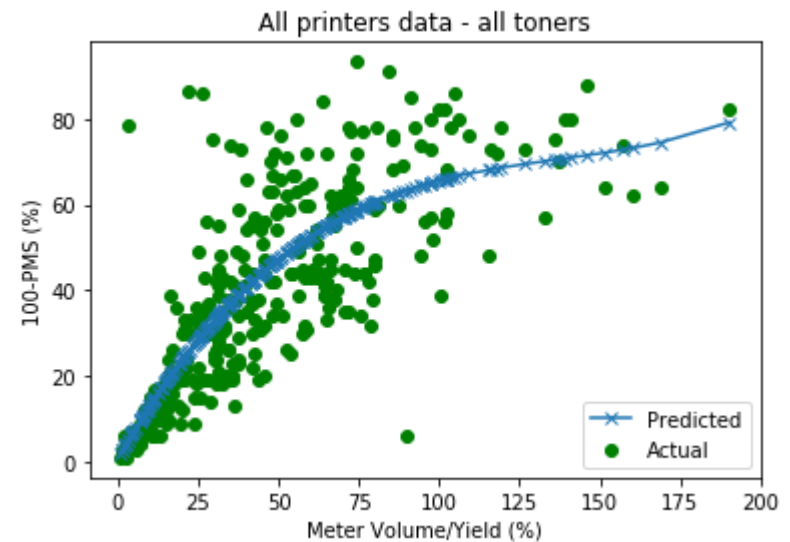
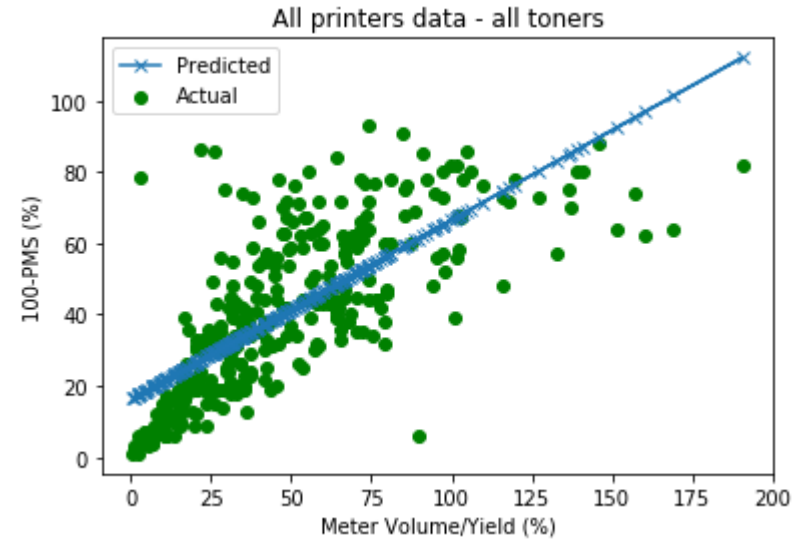
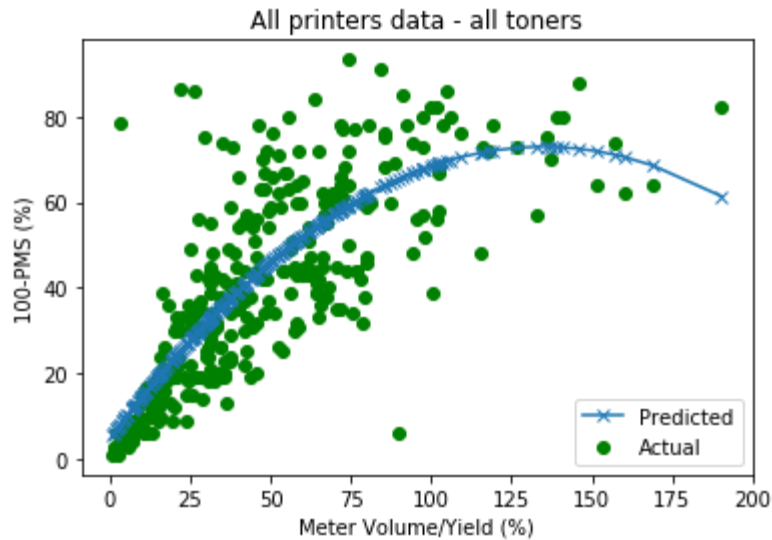
## ***Approaches to compare the actual and calculated usages***

- Single model to all printers (simple linear, quadratic and cubic models)
- Single model for each consumable type for all printers (simple linear, quadratic)
- Single model for each consumable type for each printer (simple linear)

# Analysis



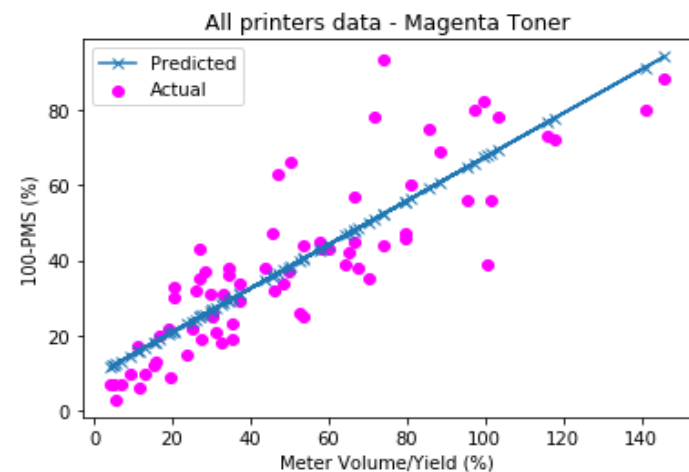
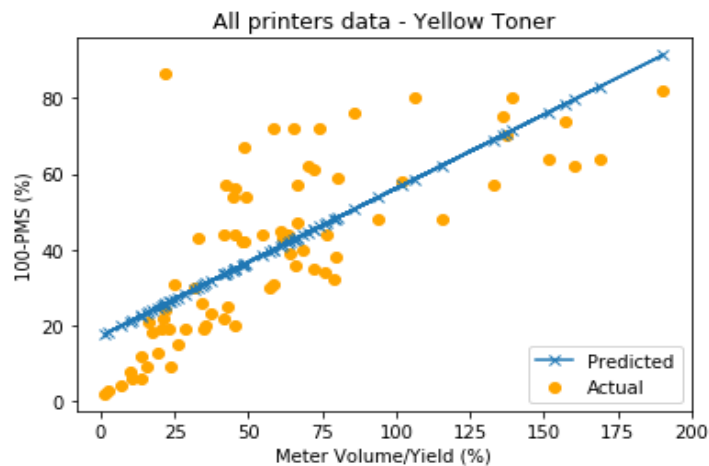
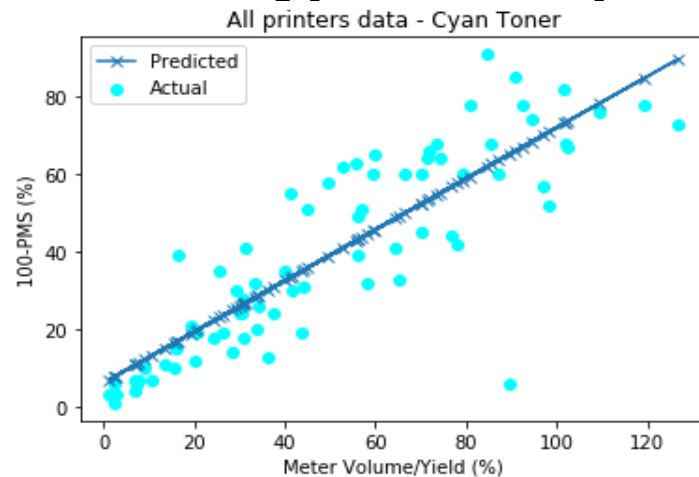
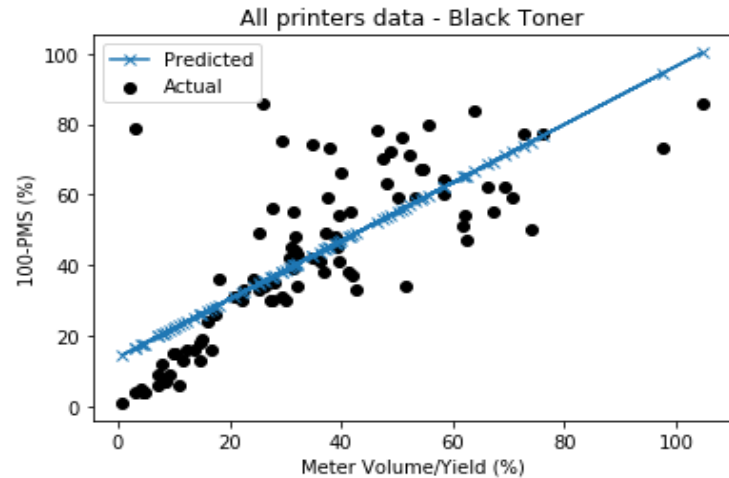
## 1. Single model to all printers



# Analysis



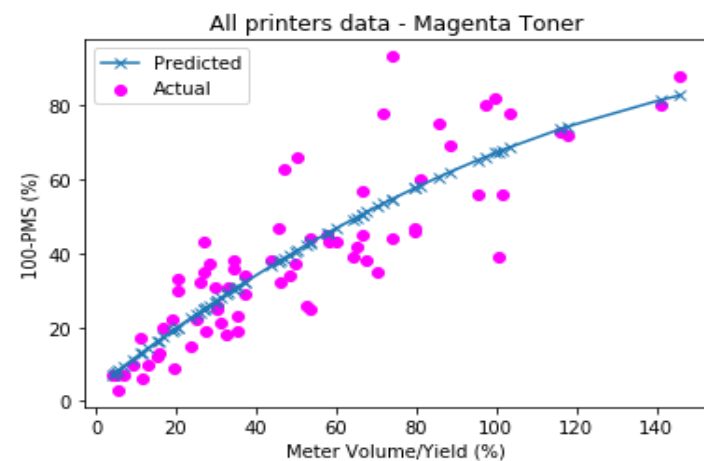
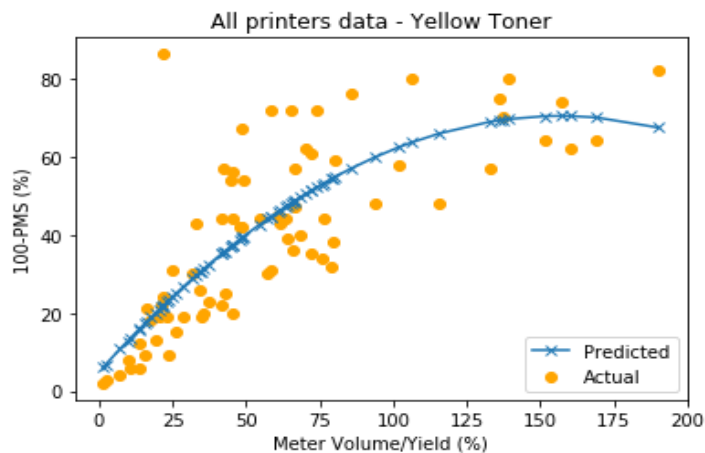
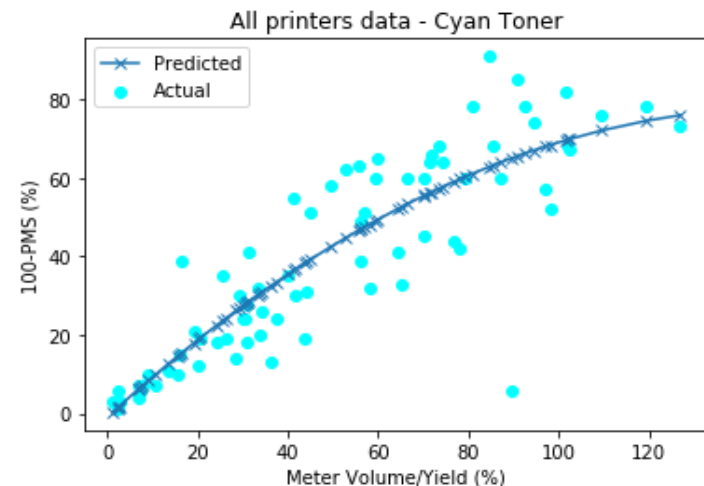
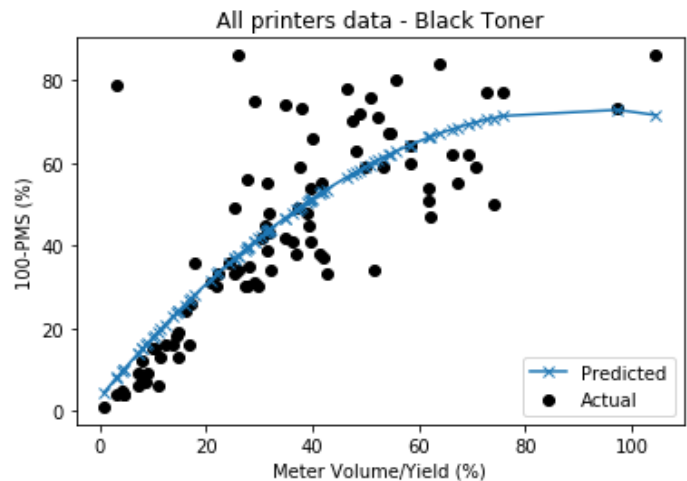
## 2. Single model for each consumable type for all printers



# Analysis



## 2. Single model for each consumable type for all printers

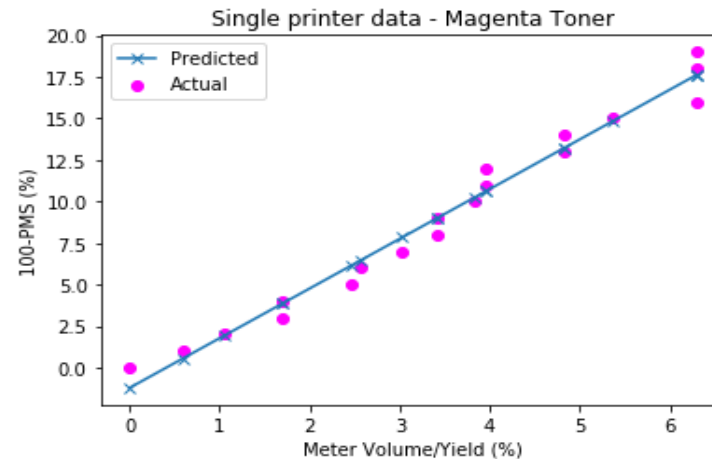
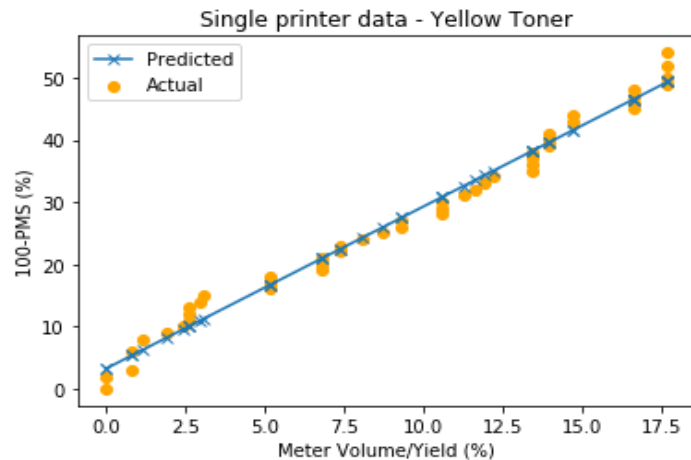
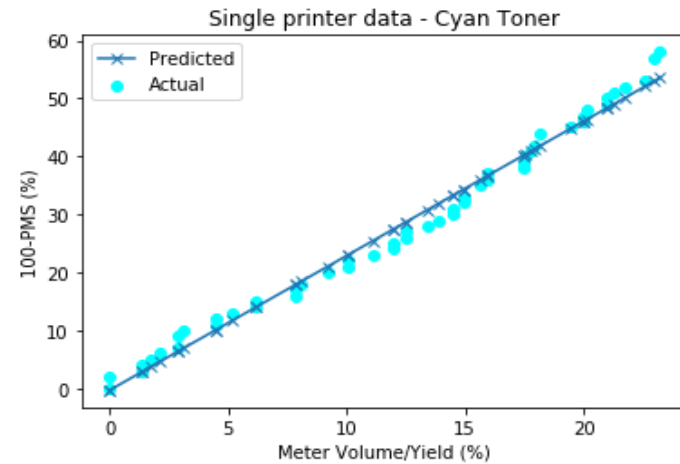
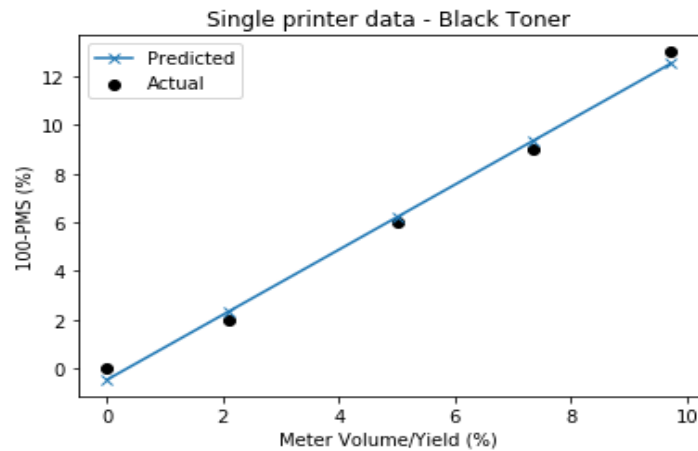




# Analysis



## 3. Single model for each consumable type for each printer



# Summary



Regression Model	Categories	Negative Mean Square (Training Set)	Negative Mean Square (Test Set)	R <sup>2</sup> (Training Set)	R <sup>2</sup> (Test Set)
Simple Linear	All printers data - all toners	252.848	259.666	0.539	0.508
Simple Linear	All printers data - Black Toner	226.093	237.786	0.587	0.519
Simple Linear	All printers data - Cyan Toner	168.136	173.31	0.729	0.681
Simple Linear	All printers data - Yellow Toner	227.165	244.384	0.55	0.407
Simple Linear	All printers data - Magenta Toner	128.1	143.65	0.747	0.699
Simple Linear	Single printer data - Black Toner	0.098	0.777	0.995	-0.11
Simple Linear	Single printer data - Cyan Toner	2.617	15.234	0.988	0.16
Simple Linear	Single printer data - Yellow Toner	2.763	6.987	0.984	0.577
Simple Linear	Single printer data - Magenta Toner	0.653	1.185	0.974	0.438
Polynomial (Quadratic)	All printers data - all toners	211.038	214.815	0.616	0.594
Polynomial (Quadratic)	All printers data - Black Toner	195.963	211.652	0.642	0.569
Polynomial (Quadratic)	All printers data - Cyan Toner	154.69	161.872	0.75	0.705
Polynomial (Quadratic)	All printers data - Yellow Toner	191.866	204.603	0.619	0.48
Polynomial (Quadratic)	All printers data - Magenta Toner	120.854	133.323	0.761	0.72
Polynomial (Cubic)	All printers data - all toners	207.67	210.464	0.622	0.602

# Summary



Regression Model	Categories	Intercept B0	Coefficients B1	Ideal Asset Coverage
Simple Linear	All printers data - all toners	16.255	0.503	2.51
Simple Linear	All printers data - Black Toner	13.895	0.825	4.13
Simple Linear	All printers data - Cyan Toner	6.22	0.658	3.29
Simple Linear	All printers data - Yellow Toner	17.252	0.39	1.94
Simple Linear	All printers data - Magenta Toner	9.298	0.581	2.903
Simple Linear	Single printer data - Black Toner	-0.461	1.336	6.68
Simple Linear	Single printer data - Cyan Toner	-0.324	2.32	11.6
Simple Linear	Single printer data - Yellow Toner	3.228	2.606	13.03
Simple Linear	Single printer data - Magenta Toner	0.653	1.185	14.959

# Conclusions and Recommendations

---

- Best to use a single simple linear regression model for each printer per consumable. This approach can help predict with a much higher accuracy the Actual usage values.
- Implementing this is easier since it already has a common data connector to all the sources of data needed to provide and calculated the daily Actual and Calculated usage levels for each printer
- It is not recommended though to use a single model for all consumables since the performance of this case is not good for all the models we tried



---

Thank you!!!