A Simple Forecasting Method for Oil Palm Plantation Yield

By: Khoo Ter Yang

Thesis Advisor: Dr Shardul S. Phadnis

ABSTRACT

A leading global player of sustainable palm oil recognised forecast accuracy of its plantation yields as crucial for making strategic and tactical supply chain operational and financial decisions. It launched a public competition in search of a superior forecasting method. In response to this call for arms, this research developed an easily explainable and implementable model for forecasting monthly oil palm yield at the granular level of individual fields (30 – 200 hectares) 12 months in advance of harvesting, with a demonstrated accuracy and consistency of 70% complying with the competition's criteria. For practical benefit of the model lies in a forecast interval scale that allows balance between inclusivity and interval width. The scholarly contribution of this research weighs big data analysis and machine learning techniques against simpler forecasting methods, which are known to outperform more complex ones.

Mr. Khoo Ter Yang received a Bachelor of Engineering (Honours) at Monash University. He has led lean six sigma management programs in two of the largest oil palm companies over a total of ten years.

KEY INSIGHTS

1. A simple forecast model based on data from preceding years can predict field-level yield with70% accuracy 12 months in advance of harvesting.
2. Simple forecasting methods outperform more complex methods in accuracy and consistency.

INTRODUCTION

In April 2019, the Head of Digital Strategy of Palm Oil Co. (*pseudonym*), one of the largest global palm oil companies, kicked off the journey towards "precision agriculture" with a public competition to identify the best method for forecasting oil palm plantation yield. Prior to this date, the company had been forecasting *annual aggregate* yields for over 6,000 fields at an accuracy of 90%. The precision agriculture competition sought forecasting at the *monthly* level, 12 months in advance of harvesting for *individual field*s with an accuracy of at least 70% (which translates to at most 30% mean absolute percentage error - MAPE). Palm Oil Co provided a set of 10 years data (2008 – 2017) with the challenge of forecasting the yield for 2018. For this study, consistency was measured as the percentage of months that achieved this accuracy target of 70% or higher. The consistency target was also set at 70%. At the time of this research, no available literature provided a satisfactory forecast model for oil palm yield at such accuracy and granularity.

Yield forecast is important in this industry for financial and operational decision making. Fresh fruit bunch (FFB) sourcing and logistics, mill utilisation, and number of harvesters depend on current yield levels but are secured based on forecasts made months in advance. FFB and crude palm oil (CPO) trading prices are negotiated and hedged based on yield forecasts as well.

EXTANT KNOWLEDGE

Many scholarly papers have demonstrated that when it comes to forecasting accuracy, simple is better. The seminal literature on this topic involves the four M-competitions that compared forecasting methods, both simple and complex (Makridakis et al, 1982; Makridakis et al, 1993; Makridakis & Hibon, 2000; Makridakis et al, 2018). In general, the four papers concluded that simpler methods were more accurate over complex methods. Despite these conclusions, many recent studies have employed more sophisticated analytical methods with the advent of big data analysis. In the field of oil palm yield forecasting, linear regressions with climatology indicators such as rainfall, wind speed, and temperature are commonly applied. These studies however have not demonstrated the accuracy and granularity required. In this research, I studied the effects of various factors with the intention of maintaining parsimony of the model for ease of explanation and application in the fields.

RESEARCH METHODOLOGY

By analysis of variance (ANOVA), the five fields displayed significantly different yields and therefore, forecasting models for each field were developed separately.

To understand the behaviour of yield, I observed their patterns for common characteristics described by Abdullah (2012): (1) trend: a gradual increase or decrease in yield over years, (2) cyclicality: a winding pattern about the trend line with unspecified periods, and (3) seasonality: a repeated pattern observed every 12 months.

Based on the yield pattern observations, I explored five methods for forecasting in this research: simplistic, moving average, multiple regression, Winters', and adjusted Winters' methods. The simplistic method was a direct use of yield from preceding years while the moving average method is an extension of this method based on multiple preceding years. I developed moving average models based on the average yields of similar months from three and six preceding years of data. The multiple regression method forecasts yield as a linear function of historical yield and three of the highest correlated factors.

Using the Winters' method, two forecast models were generated based on yield of three and six preceding years using Minitab 2019. The Winters' model was also adjusted for error reduction by multiple regressions with the correlated factors.

RESULTS

The yearly and monthly yield of each field were plotted using boxplots to observe for patterns and variability. An example of this plot is shown for *Field 1* in **Figure 1**. The following observations on yield behaviour were similar for all five fields studied:

- In Figure 1(a), the horizontal centreline indicates no upward or downward trend in yield. This is supported statistically by a run chart test on Minitab 2019.
- The dotted-red line joining yearly median yields in a winding pattern about the centreline in indicate Cyclicality.
- The dotted-blue line joining monthly median in an s-shape curve in Figure 1(b) indicate seasonality of monthly yield.
- The unsystematic variations in the yearly and monthly yield in both Figures 1 (a) and (b) were identified as irregularities.

For monthly forecast modelling, the pattern for seasonality was most relevant as it provided predictability of monthly yield level befitting the s-shape curve each year.

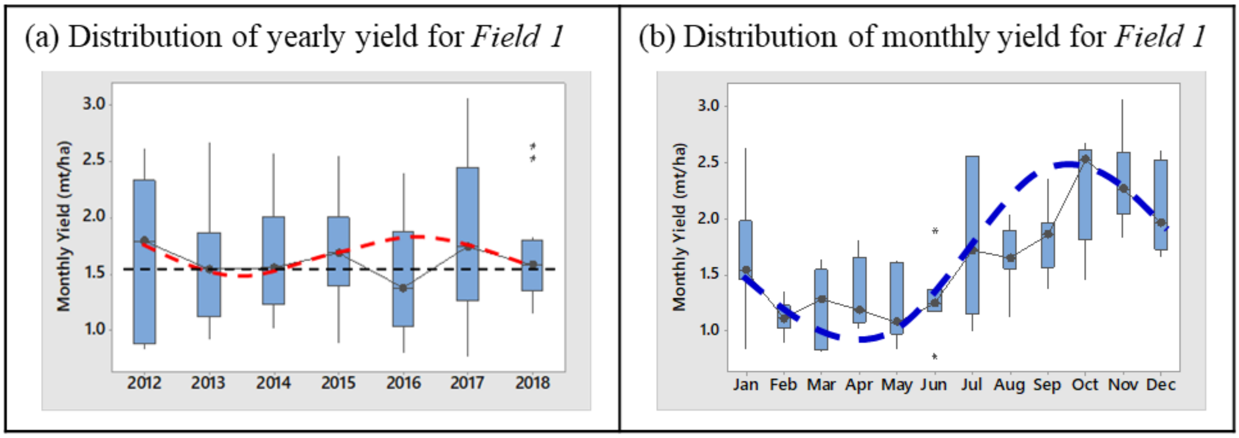

Figure 1: Distribution of yearly and monthly yield for *Field 1*.

Cyclicality and irregularity were less useful as they do not correlate with monthly yield.

The key results of this research are summarised in **Figure 2**, which compares the average accuracy and consistency of the methods studied across three years for five fields. Overall, the moving average and simplistic methods based solely on historical yield outperformed Winters' method, while the multiple linear regression method resulted in the lowest accuracy and consistency.

Across all fields in general, the moving average method based on three preceding years of yield showed the highest average accuracy (71%) and consistency (70%), both of which meet the target of Palm Oil Co and this research. Interestingly, the two methods which incorporated yield factors as predictors (multiple regression and

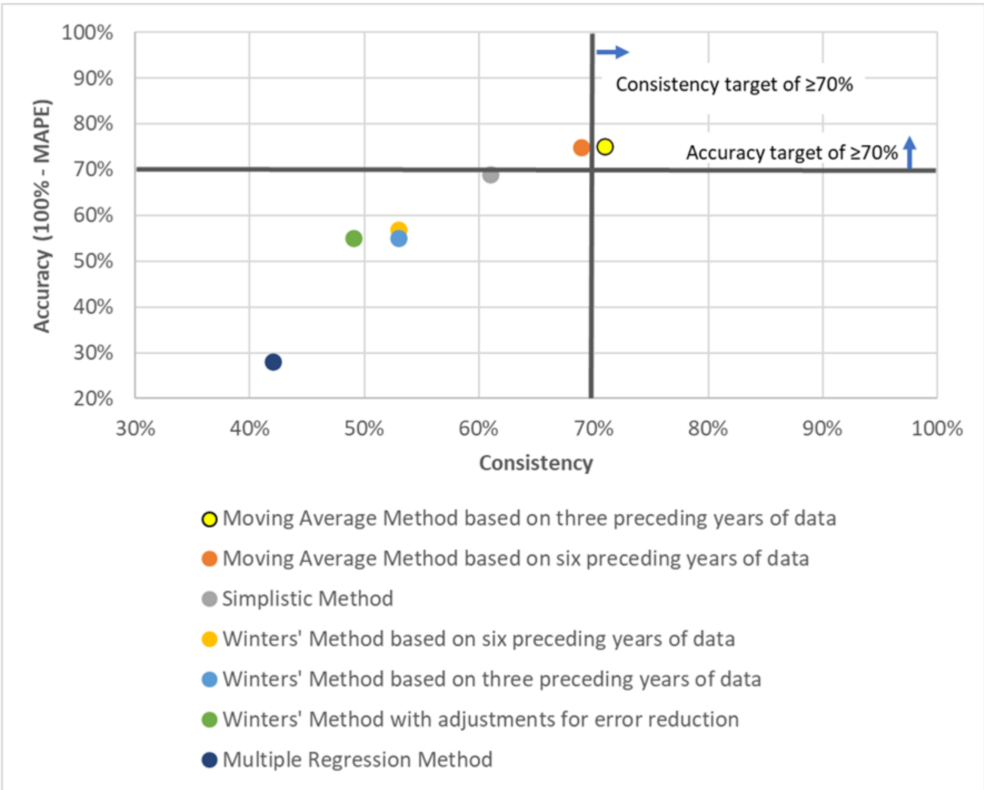

Figure 2: Comparison of average accuracy and consistency of each model.

Winters' method with adjustments for error reduction) displayed the lowest accuracy. It is highly likely that both methods had overfitted data from preceding years into the model, causing larger inaccuracies.

At a field level, the forecast models display varying accuracy and consistency. This may be due to the unique behaviour of each field as evident in the ANOVA test. Therefore, higher forecast accuracies may be achieved by selection of the best performing method for each field.

As a general approach for a forecasting method to apply across multiple fields, the moving average method based on three preceding years of data is suggested not only based on its overall accuracy and consistency across fields and years, but also because it can be easily applied and explained.

CONCLUSION

This research concurs with the robust findings in the forecasting literature that simplicity is beauty! The best performance was achieved by the relatively simpler moving average method.

This research ends with a highlight of how these findings may be applied in Palm Oil Co and areas of future work for enhancement and expansion of application.

REFERENCES

Abdullah, R. (2012). An Analysis of Crude Palm Oil Production in Malaysia. Kuala Lumpur: Malaysian Palm Oil Board.

Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. International Journal of Forecasting, 451-476.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., . . . Winkler, R. (1982). The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition. Journal of Forecasting, 111-153.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., M. T., O. K., & Simmons, L. F. (1993). The M2-Competition: A real-time judgmentally based foreasting study. International Journal of Forecasting, 5-22.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. International Journal of Forecasting.