

A Hybrid Approach to Modelling the Climate Change Effects on Malaysia's Oil Palm Yield at the Regional Scale

Subana Shanmuganathan^{1 2}, Ajit Narayanan^{1 3}, Maryati Mohamed^{4 6},
Rosziati Ibrahim^{5 6}, Haron Khalid⁷,

¹ Auckland University of Technology (AUT), New Zealand.

² Geoinformatics Research Centre (GRC),

³ School of Computer and Mathematical Sciences (SCMS)

⁴ Department of Technology and Human Development, Faculty of Science,

⁵ Research, Innovation, Commercialisation & Consultancy Office (RICCO),

⁶ Universiti Tun Hussein Onn Malaysia (UTHM), Johor, Malaysia.

⁷ MPOB Research Station Kluang, Malaysia

{subana.shanmuganathan, ajit.narayanan}@aut.ac.nz

{maryati, rosziati}@uthm.edu.my, khalid@mpob.gov.my

Abstract. Understanding the climate change effects on local crops is vital for adapting new cultivation practices and assuring world food security. Given the volume of palm oil produced in Malaysia, climate change effects on oil palm phenology and fruit production have greater implications at both local and international scenes. In this context, the paper looks at analysing the recent climate change effects on oil palm yield within a five year period (2007-2011) at the regional scale. The hybrid approach of data mining techniques (association rules) and statistical analyses (regression) used in this research reveal new insights on the effects of climate change on oil palm yield within this small data set insufficient for conventional analyses on their own.

Keywords: Regression test, Data mining (association rules), WEKA, JRip

1 Introduction

Climate change has the potential to impact on almost all forms of agricultural systems and their productivity at varying degrees e.g., from small benefits to extensive drastic effects depending on the current climate regime and the crop [1]. Understanding this variability in climate change and its effects on local crops is vital for adapting new cultivation practices to overcome the harsh effects on crops, especially to assure world food security now and in the future. Malaysia's oil palm exports contribute to 45% of the world's edible oil needs [2]. Hence, it is essential that we understand the climate change and its potential effects under different scenarios for palm oil production. The paper looks at the correlations between recent climate change and oil palm yield in Malaysia's administrative regions over a period of five years (2007-2011) to determine the trends and possible climate change effects. The work presented here is

an extension to earlier work that looked at possible climate change effects in West (Peninsular) and East (Borneo) Malaysia, outlined in section 2.3 of this paper.

2 Previous research

Previous research results provide clues on the phenological stages that are susceptible to the extreme climate events and are outlined in this section.

2.1 Effects of El Niño and La Niña in Tumaco, Colombia

In a pioneering study [3], the effects of El Niño and La Niña events on the local climate and on oil palm tree (*Elaeis guineensis*) production in the Tumaco situated in the Pacific Coast of Colombia were reported. In that study, it was established that El Niño and La Niña events had lagged (previous period) and conflicting impacts on oil palm yields. El Niño experienced in 1997/98 was found to be favourable showing the maximum correlation with production 2.6 years after the event. Meanwhile, La Niña of 1999/2000 had caused severe droughts, with the highest negative impact (reduced yield) in 2002.

2.2 Effects of El Niño Events in Sabah, Malaysia

In [4] using statistical analyses, the authors established that heavy rains and high temperatures as favourable to palm oil production in the western coast of Sabah with a lag period of 3 and 4 months respectively. However, the extremes of both scenarios, i.e. flooding and severe drought, were found to be not necessarily favourable to oil palm production. The extremes of La Niña events, e.g. higher precipitation/floods, decreased the production and quality of crude palm oil (CPO), the events being attributed to affecting the fruit ripening stage (fig. 6) reflected in the yield at later months. The Sabah study [4] looked at the climate variability in rainfall, temperature and their correlations with fresh fruit bunch (FFB) production and fish landings in the West coast of Sabah, Malaysia, using data from 2000 to 2010.

2.3 Climate Change Effects on Oil Palm in Malaysia

In view of the above two studies in [5] an attempt was made to model the climate change variability and its effects on oil palm yield in the East (Peninsular) and West (Borneo/ Sabah and Sarawak) Malaysia located on either side of the South China Sea. The results of that initial research using monthly climate anomalies in temperature and oil palm yield from the two main regions of Malaysia revealed new insights on the correlations between the climate variations, oil palm tree phenology and yield. The climate variability in the two main Malaysian regions and the lagged effects of temperature on the oil palm yield conformed to the phenological growth stages of the oil palm tree, established as vulnerable to climate change in earlier work in the

Columbian and Malaysian contexts (sections 2.1 and 2.2 respectively). However, it was found that further research was required to establish the precise effects of temperature and the phenological stages vulnerable to temperature, using long term yield and weather data especially at the regional scale.

The aim of this paper is to extend our growing knowledge of lag periods and their effects on palm oil growth and production. In particular, there are still questions on how changes in lag period suggested in the earlier studies, are related to production yield. Such knowledge is critical for economic and production planning purposes so that, after a severe climate event, appropriate plans and actions can be put in place to limit the effects of the severe event at suitable time points subsequently.

In the next section the methodology and data used in this research are discussed.

3 The Methodology

For this research, monthly averages of temperature and oil palm yield (Fresh Fruit Bunch/ FFB) data at the regional scale were used for analysing the correlations between the climate and yield at this scale. The data set is analysed using a hybrid approach consisting of Top-Down Induction of Decision Trees (e.g. JRip function in WEKA software) and statistical regressions. Two sets of regressions were run, firstly, one test with all regional yield and lag variables together and then 10 more regressions were run with each of the individual regional data separately. The WEKA rules (generated for all regions) and regression results of both all regions together and the 10 individual regions alone were compared to investigate the trends in the climate change effects on oil palm yield across Malaysia in general and those specific to the individual regions.

3.1. Oil Palm Yield Data

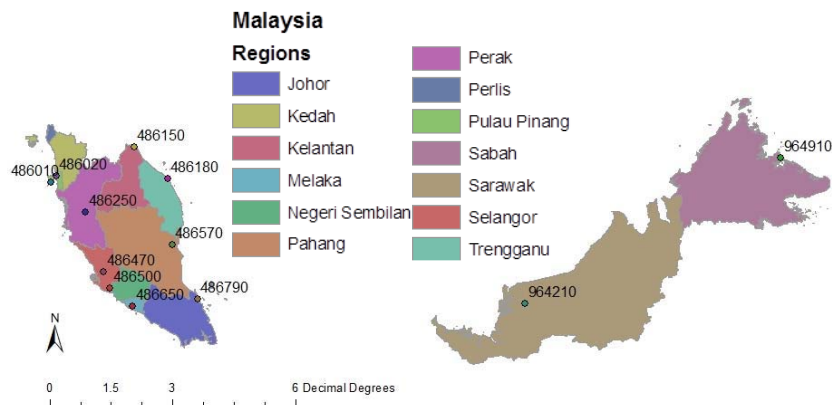


Fig. 1. Map of Malaysia's administrative regions and weather stations (table 1) [7].

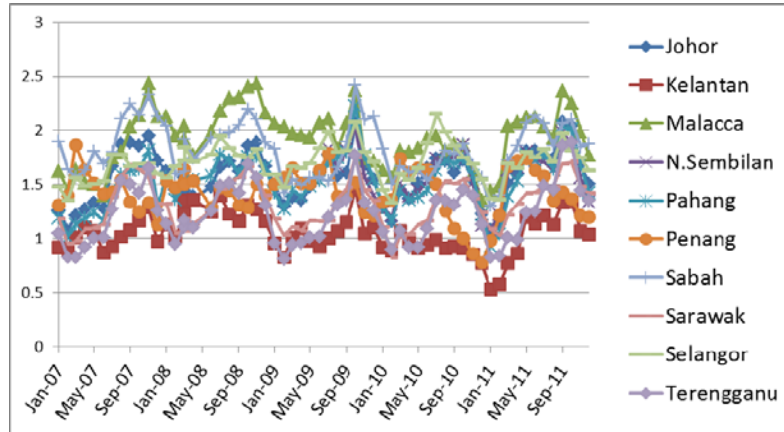


Fig. 2. Graphs showing oil palm yield (fresh fruit bunch/FFB in T/H/M) in the ten Malaysian regions studied in the research.

Table 1 (left). Yield regions and weather stations (global daily summary temperature) data was obtained. Source: NCDC Climate Services Branch (CSB) [7].

Sta no.	Station name	Yield region
486010	PENANG/BAYAN LEPAS	Penang
486020	BUTTERWORTH	Kedah
486150	KOTA BHARU	Kelantan
486180	KUALA TRENGGANU	Terengganu
486250	IPOH	Perak
486470	KL SUBANG	Selangor
486500	KUALA LUMPUR INTL N.	Sembilan
486570	KUANTAN	Pahang
486650	MALACCA	Malacca
486790	JOHORE BHARU/SENAI	Johor
964210	SIBU	Sabah
964910	SANDAKAN	Sarawak

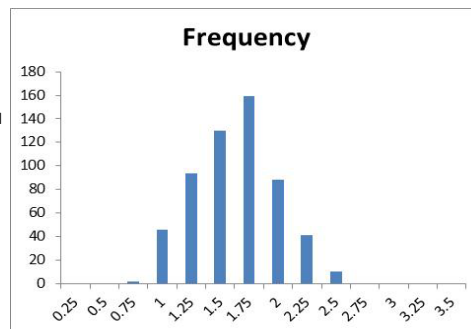


Fig. 3 (right). Oil palm yield frequency distribution within 2007-2011 monthly regional yield data. WEKA yield classes (FFB in T/H/M) Low <1.25 , Medium >1.25 <2 , and High >2

Monthly oil palm yield (FFB) data for the regions (fig. 1) for a period of five years (2007-2011) obtained from [6], presented as graphs (fig. 2) was analysed against 36 lag variables (36 monthly average temperatures prior to harvest).

3.2 Yield Classes

Based on the data distribution and frequency histogram (fig. 3), yield (FFB in T/H/M) was classified into three classes for analysis using supervised learning in WEKA: Low <1.25 , Medium /Med ≥ 1.25 <2 , and High ≥ 2 .

3.3. Climate Data

Global daily temperature summaries extracted from [7] for stations (fig 1 and table 1) in the respective regions were converted into a table of monthly average temperatures (°C) for the corresponding 10 regions of Malaysia studied in this research.

In this research, 36 monthly temperature averages calculated for the 10 regional stations (fig. 1 and Table 1), and the monthly oil palm yield for these individual regions (fig. 2) were analysed to determine the lag effects of the climate on oil palm yield across Malaysia at the regional scale and the results are presented in next section

4 The results

The variability in regional monthly and annual yield appears to have some common patterns across Malaysia and some other specific characteristics unique to the individual regions (figs. 2 and 4). For example, the yield is high mostly in the north western regions of Peninsular Malaysia and in Sabah and this can be observed throughout the study period 2007-2011 (fig 4).

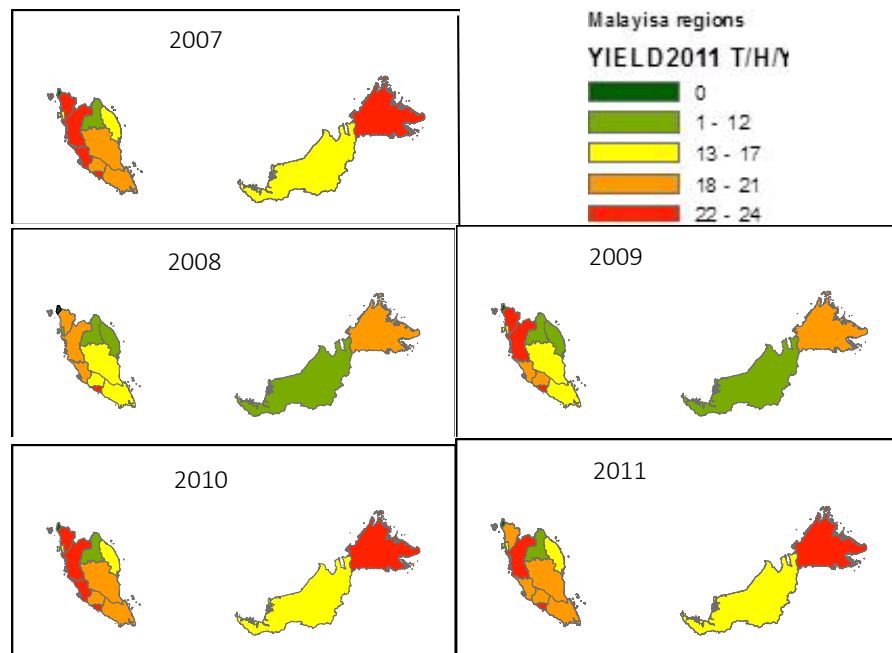


Fig. 4. Annual Oil Palm yield (T/H/Y) variability during 2007-11.

4.1. WEKA Results

The WEKA JRip rule function run with 36 monthly average temperatures prior to harvest (lag variables relating to berry development cycle) and the monthly yield of all 10 regions produced association rules that classified the yield classes at 78.9% accuracy (fig 5). The 7 rules (fig. 5) produced by the WEKA function showed the critical monthly temperatures that had had lagged effects on the yield in five specific regions (Malacca, Sabah, Kelantan, Sarawak and Terengganu), as well as across Malaysia.

JRIP rules:							
<hr/>							
1. (state = <u>Malacca</u>) and (lag7 <= 27.55) and (lag10 <= 27.35)							
=> yield class=high (14.0/1.0)							
2. (state = <u>Sabah</u>) and (lag28 >= 26.86) => yield class=high (13.0/4.0)							
3. (state = <u>Kelantan</u>) => yield class=low (60.0/10.0)							
4. (lag33 >= 27.53) and (state = <u>Sarawak</u>) => yield class=low (28.0/6.0)							
5. (state = <u>Terengganu</u>) and (lag2 <= 27.28) => yield class=low (23.0/2.0)							
6. (lag8 >= 27.83) and (lag23 <= 26.62) => yield class=low (16.0/5.0)							
7. (lag21 >= 27.29) and (state = <u>Terengganu</u>) and (lag16 <= 26.77)							
=> yield class=low (7.0/0.0)							
8. => yield class=med (409.0/56.0)							
<u>Number of Rules: 8</u>							
<u>Time taken to build model: 0.32 seconds</u>							
=== Stratified cross-validation ===							
=== Summary ===							
Correctly Classified Instances	450						78.9474 %
Incorrectly Classified Instances	120						21.0526 %
Kappa statistic	0.5441						
Mean absolute error	0.1918						
Root mean squared error	0.3408						
Relative absolute error	59.0095 %						
Root relative squared error	84.616 %						
Total Number of Instances 570							
=== Detailed Accuracy By Class ===							
TP	Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.882	0.397	0.818	0.882	0.848	0.755	med
	0.674	0.056	0.795	0.674	0.729	0.836	low
	0.412	0.04	0.5	0.412	0.452	0.747	high
WA.	0.789	0.282	0.784	0.789	0.784	0.774	
=== Confusion Matrix ===							
a	b	c <-- classified as					
336	24	21 a = med					
45	93	0 b = low					
30	0	21 c = high					

Fig. 5. JRip (WEKA rules) generated using yield classes (low, medium and high) as dependent and lag variables (36 months prior to harvest) as independent factors.

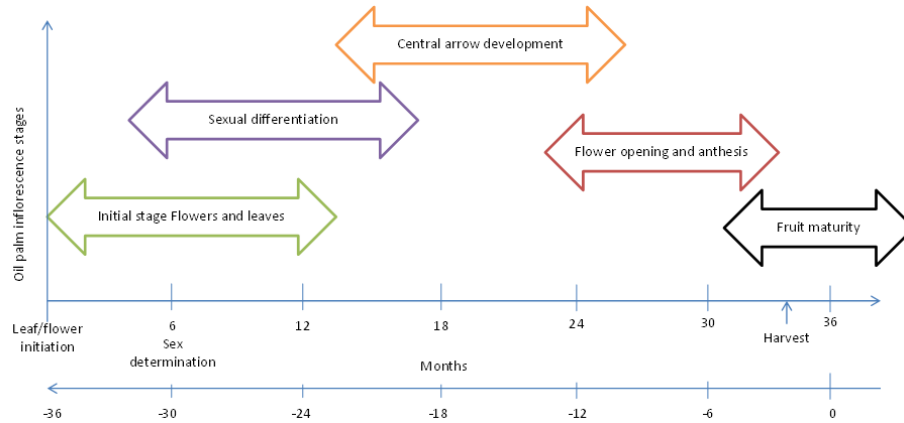


Fig. 6. Oil palm life cycle (frond emergence month 1/36 to -36/0 to fruit harvest) with stages and approximate timing relating to bunch development.

The first two JRip rules (fig. 5) showed the lagged effects of temperature that had led to “high” (>2 T/H/M) yield in Malacca (14 instances/ 1 exception) and Sabah (13 instances/ 4 exceptions). For the Malacca region, temperatures in month -7 ($\leq 27.55^{\circ}\text{C}$) and -10 ($\leq 27.35^{\circ}\text{C}$) had led to “high” yield. Similarly, for Sabah temperature in month -28 ($\geq 26.86^{\circ}\text{C}$) had led to “high” yield. In oil palm phenology months -7 to -10 are the time when flower opening and anthesis occur, and month -28 relates to the initial frond emergence and sex determination in the berry development cycle (fig. 6).

Rule 3 (fig. 5 with 60 instances/10 exceptions) implies that the yield in Kelatan is “low” when compared with that of the other regions. Based on rule 4, in Sarawak, lag variable/ month -33 $\geq 27.53^{\circ}\text{C}$ temperature had led to “low” yield, and this relates to the frond emergence phase (fig. 6).

Based on rule 5, in Terengganu, lag variable, month -2 $\leq 27.28^{\circ}\text{C}$ had led to “low” yield, and this relates to the fruit maturity period (fig. 6).

Rule 6 for all regions relates to lag variable month -8 $\geq 27.83^{\circ}\text{C}$ and month -23 $\leq 26.62^{\circ}\text{C}$ that had led to “low” yield with 16 instances and 5 exceptions. This implies that higher temperatures during anthesis/ early fruit development and lower temperatures during frond /leaf emergence had led to “low” yield across the country.

Based on rule 7, in Terengganu lag variable month -21 $\geq 27.29^{\circ}\text{C}$ and month -16 $\leq 26.77^{\circ}\text{C}$ had led to “low” yield, met in 7 instances. Hence, in Terengganu also, high/ low temperatures during flower opening and sex determination stage had affected on the region’s yield (fig. 6).

4.2 Regression test results

From the first regression test run on all 10 regional data together, month -9, -13 and -8 temperatures were shown as the most “critical lag variables” or predictors for the monthly yield (fig. 7). Earlier studies from literature [2][4] and authors of this paper [5] established the phenological stages relating to these months as: flower opening and anthesis (month -8/-9), and sexual determination month -13 (fig. 6).

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	5.904	.482		12.252
	lag9	-.162	.018	-.357	.000
2	(Constant)	3.942	.644		6.121
	lag9	-.171	.018	-.377	.000
	lag13	.081	.018	.175	.000
3	(Constant)	4.502	.676		6.658
	lag9	-.117	.027	-.257	.000
	lag13	.077	.018	.165	.000
	lag8	-.071	.027	-.155	.010

a. Dependent Variable: yield

ANOVA ^a					
Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	9.066	1	9.066	83.154	.000 ^a
Residual	61.930	568	.109		
Total	70.996	569			
2 Regression	11.204	2	5.602	53.122	.000 ^b
Residual	59.792	567	.105		
Total	70.996	569			
3 Regression	11.907	3	3.969	38.016	.000 ^c
Residual	59.090	566	.104		
Total	70.996	569			

a. Predictors: (Constant), lag9
b. Predictors: (Constant), lag9, lag13
c. Predictors: (Constant), lag9, lag13, lag8
d. Dependent Variable: yield

Model Summary ^d									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.357 ^a	.128	.126	.3301990	.128	83.154	1	568	.000
2	.397 ^b	.158	.155	.3247367	.030	20.269	1	567	.000
3	.410 ^c	.168	.163	.3231077	.010	6.732	1	566	.010

a. Predictors: (Constant), lag9
b. Predictors: (Constant), lag9, lag13
c. Predictors: (Constant), lag9, lag13, lag8
d. Dependent Variable: yield

Fig. 7. The regression results of all 10 regional yields and their lag variables run together. The predictors (lag variables) average temperatures of month -9, -13 and -8 correspond to flower opening and anthesis of oil palm phenology (fig. 6)

The 10 different regressions run on 36 lag variables against the respective regional yield individually produced the respective predictor/s (lag variables/ monthly average temperatures) for the yield at the regional scale (table 2). In the regions studied, month -7 or -8 or -9 or a combination of these monthly temperatures were shown to be having lagged effects on the yield and this -7 to -9 month (or prior to harvest) period corresponds to flower opening and anthesis stage (fig. 6).

For Kelantan, N. Sembilan, Pahang, Malacca, Johor and Sabah the critical temperature is in month -9. Hence, this is one of the predictors for yield. These regions are found in the southern Peninsula Malaysia and northern Boneo Malaysia except for Kelantan on the north east tip. Similarly, for Selangor and Penang it is month -8. Sarawak -10 and for Terengganu -7 are the months found to be among the predictors. Malacca has -9 as the sole predictor for the region’s yield, and again this is the period flower opening and anthesis occur. It seems that in all the regions temperature affects the flower opening and anthesis, with their lagged effects reflected on each of the region’s yield in 7-10 months later.

Table 2. The regions, months and R² that showed lagged effects on yield. (-27 indicates the model without the lag variable average temperature of month 27.

Station No	Region	lag variable/ month-	R square
486010	Penang	23, 27, 6, 8, 31	0.43 0.50 0.59 0.59 0.63 0.71
486020	Kedah	---	
486150	Kelantan	9, 4	0.28 0.39
486180	Terengganu	33, 27, 21, 7, 15, 9	0.45 0.63 0.69 0.73 0.75 0.78 0.78 (-27
486250	Perak	---	
486470	Selangor	8, 29, 20, 33, 17 15	0.30 0.49 0.59 0.63 0.65 0.68
486500	N. Sembilan	9, 31, 35, 1, 3	0.45 0.60 0.67 0.72 0.79
486570	Pahang	9, 13, 29, 33, 25	0.51 0.67 0.72 0.75 0.79
486650	Malacca	9,	0.27
486790	Johore	9, 2, 29, 13	0.33 0.47 0.66 0.73
964210	Sabah	9, 14, 17, 13, 31	0.38 0.46 0.54 0.60 0.62
964910	Sarawak	33, 27, 13, 29, 7, 10, 14	0.50 0.66 0.72 0.77 0.80 0.82 0.84 0.84
			(-27

The average temperatures of month 29-35 prior to harvest as well showed lagged effects on the yield in 8 of the 10 regions (table 2), which is the frond/leaf emergence stage, also confirmed by JRip rule 6 (fig 6). The other stage shown as critical in the JRip rule 6 is the month -8, which is the flower anthesis phase in the oil palm berry bunch development cycle, also confirmed by all regressions (all regional data together and individual regions separately). Meanwhile, in Sarawak and Terengganu regions, frond emergence stage (month -33) seems be the primary predictor for yield.

The regressions run on individual regional data separately gave higher adjusted R square values for the models, 0.27 to maximum of 0.84 as opposed to 0.128-0.168 in the all regions together. This shows that the influence exerted by the lag variables on the yield varies for the different regions but for the same phenological phase (flower opening and anthesis at 7-13 months prior to harvest) hence, the effects cannot be generalised across regions. Further analysis incorporating local data (e.g. rainfall, terrain) could shed further insights on the precise climate change effects on yield.

Conclusions

The paper looked at the possible climate change effects on Malaysia's oil palm yield using 36 monthly average temperatures as lag variables along with yield data at the regional scale. Even with this small data set from 2007 to 2011, this initial investigation conducted at this scale has provided some useful insights on the variability in the lag effects of temperature on the monthly oil palm yield. For instance, based on JRip rule 6 lag variable, average monthly temperature -8 (8 months prior to harvest) $\geq 27.83^{\circ}\text{C}$ has led to "low" yield (<1.25 T/H/M) across all regions of Malaysia.

Rules 2 and 6 indicate that subtle increases e.g., $27-28^{\circ}\text{C}$ in monthly average temperatures at the initial stages of fruit development, have led to "high" yields in Sabah.

From the regression run on all regional yields together, lag variables average temperatures of month -9, -13 and -8 (flower opening and anthesis) were found to be the predictors. This has also been reflected in the individual regional regressions but the predictability of the lag variables were much higher. This implies that in all the regions the effects of temperature in the early stages of the fruit development are lagged and reflected in the yield in 8-13 months later.

The adjusted R square value for all 10 regions together regression model (month -9, -13 and -8) was low (0.128-0.168 as opposed to the 0.27-0.84 of the individual regional regression models) and this indicates that the effects of climate change on oil palm yield for the whole nation could not be generalised.

In conclusion, this research indicates that attempting to predict the effects of climate change on palm oil yield in particular, and perhaps other crops more generally, will need to be regionalised if accuracy of prediction is required. The exact size of region will depend on local environmental conditions. Further research at finer scales, such as regional or blocks that are topographically discrete, is warranted to establish the climate change effects at the “*meso*” and micro scales (among and within oil palm plantations).

References

1. Iglesias, A, and C Rosenzweig. 2009. Effects of Climate Change on Global Food Production, Special Report on Emission Scenarios.
<http://sedac.ciesin.columbia.edu/mva/cropclimate/>
2. USDA-FAS Office of Global Analysis, “Indonesia: Palm Oil Production Prospects Continue to Grow www.pecad.fas.usda.gov/highlights/2007/12/Indonesia_palmoil/.”
3. M. C. Cadena, A. Devis-Morales, J. D. Pabón, I. Málikov, J. A. Reyna-Moreno and J. R. Ortiz, Relationship between the 1997/98 El Niño and 1999/2001 La Niña events and oil palm tree production in Tumaco, Southwestern Colombia, Adv. Geosci., 6 pp. 195-199 2006
4. P. P. Wen and M. J. Sidik, “Impacts Of Rainfall, Temperature And Recent El Niños On Fisheries And Agricultural Products in The West Coast Of Sabah 2000-10,” BORNEO SCIENCE 28: MARCH 2011, pp. 73-85, 2011.
5. Shanmuganathan S and Narayanan A, Modelling the climate change effects on Malaysia’s oil palm yield in proceedings of 2012 IEEE Symposium on IS3e 2012, Kuala Lumpur, Malaysia, 21-24 Oct 2012
6. Malaysia Oil Palm Board, “Statistics-Yield,” Economics and Industry Development Division, Kelana Jaya Selangor
7. NOAA, 2012 www.ncdc.noaa.gov/cdo-web/