Computers and Electronics in Agriculture 151 (2018) 338–348

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Using Bayesian networks to predict future yield functions with data from commercial oil palm plantations: A proof of concept analysis

Ross Chapman[a], Simon Cook[c,d], Christopher Donough[b], Ya Li Lim[b], Philip Vun Vui Ho[e], Koon Wai Lo[e], Thomas Oberthür[b,*]

[a] 9 Kia Ora Parade, Ferntree Gully, Victoria 3156, Australia
[b] IPNI Southeast Asia Program, 29C-03-08 Maritime Piazza, Karpal Singh Drive, Penang 11600, Malaysia
[c] The Centre for Crop and Disease Management, Curtin University, Kent Street, Bentley, WA 6102, Australia
[d] Murdoch University, 90 South Street, Murdoch, Western Australia 6150, Australia
[e] Wilmar International Group Plantations, Multivision Tower Lt. 12, Jakarta Selatan, DKI Jakarta 12980, Indonesia

ARTICLE INFO

Keywords:
Oil palm
Estate management
Machine learning
Pattern recognition
Big data
Bayesian networks
Artificial neural networks
Yield
Fertiliser
Rainfall
Predictions
Forecasts

ABSTRACT

Bayesian networks were used to predict yield functions from three commercial oil palm estates. The networks were trained using a range of environmental, agronomic and management data routinely collected during plantation management. The Bayesian networks predicted fruit yield (FFB), average weight of fruit bunches (ABW) and average bunch number per hectare (BUNCH_HA). Comparing the predictions of most probable yield against observed data showed the Bayesian networks were highly accurate, with r^2 values between 0.6 and 0.9. Predictions for attaining specific yield targets exceeded 75% accuracy for the FFB, 85% for the BUNCH_HA, and 90% for the ABW function. Supplementary analysis compared the precision of the Bayesian networks with artificial neural networks (ANNs), and demonstrated that the Bayesian networks gave equivalent or superior accuracy for every test. The utility of the networks were demonstrated by predicting the probability of achieving above average yield functions for each block across the three estates using a set of hypothetical rainfall and fertiliser input scenarios during the year prior to harvest. For the majority of blocks, the probability of exceeding the yield target depended on the level of fertiliser and rainfall inputs received, indicating that production from these blocks is greatly influenced by prior rainfall and fertilizer. However, some blocks in favourable areas showed a very high probability of exceeding the mean yields at all rainfall and fertiliser inputs, while a number of other blocks showed a consistently low probability of achieving the same productivity; production from these blocks will be resistant to the effects of historic rainfall and fertiliser inputs. The ability of Bayesian networks to represent future yield expectations will greatly assist managers under pressure to improve the economic and environmental sustainability of plantations. The demonstration that machine learning can extract important insight from complex datasets will have broad application in the analysis of big data collected from oil palm as well as other agricultural industries.

1. Introduction

The global oil palm industry has grown rapidly over recently, with production increasing from 17.64 million tonnes in 1996/97 up to 69.77 million tonnes in 2016/17 (USDA, 2018). However, the oil palm industry is facing mounting environmental, economic and political pressures which endanger future sustainability (Carlson et al., 2013). The industry's on-going resilience and profitability will depend on the ability of estate managers to make strategic and process orientated adaptations to management (Cook et al., 2014).

Plantation managers are under intense pressure to make rapid management decisions about many issues, from personnel to strategy; from area to input. Decisions are frequently made under duress and based on intuition, which often gives a sub-optimal outcome. Furthermore, managers might attach false confidence to their intuition, leading to impulsive decisions that are untested against data. The potential and cumulative risks are grave.

Decision support systems can assist managers by summarising data driven analysis and providing objective and rational perspectives of complex production systems. For example, PALMSIM is a computer simulation model that has been developed for oil palm (Hoffmann et al., 2014). However, yield predictions from this model are based solely

* Corresponding author.
E-mail address: TOberthur@ipni.net (T. Oberthür).

https://doi.org/10.1016/j.compag.2018.06.006
Received 20 October 2017; Received in revised form 31 May 2018; Accepted 2 June 2018
Available online 18 June 2018
0168-1699/

upon current solar radiation, water availability and tree age, and so are unable to represent the variation relating to environmental or management parameters. The development of a comprehensive computer model for oil palm presents many challenges. First, parameter quantification is costly and time consuming. Second, it can be difficult to generalise between contrasting geographic and environmental locations. Third, the output is typically a simple single predicted yield arising from specified environmental and management variables.

Recently, "big data" has become an increasingly common paradigm across many domains, including agricultural research (Kamilaris et al., 2017). Big data typically represents extremely large data collections characterised by the 5 Vs: **V**olume, **V**elocity, **V**ariety, **V**eracity and **V**alorisation (Chi et al., 2016; Kamilaris et al., 2017), which describe the quantity of data; the time window over which the data is relevant; the diversity of data types and sources; the quality, accuracy and reliability of the data; and the ability to propagate knowledge and innovation.

Efficient exploitation of the emerging agricultural big data resources has been estimated to offer an annual global benefit of up to $20 billion, yet, despite this potential, analysis of big data in agriculture has lagged behind other industries (Kamilaris and Preafeta-Boldu, 2018). The raw data itself presents little if any economic value, but must first be transformed into high-value knowledge and wisdom using appropriate analytical tools aligned with the Data-Information-Knowledge-Wisdom hierarchy (Rowley, 2007; Lokers et al., 2018) which can in turn be used to construct actionable management (Antle et al., 2017; Morota et al., 2018).

Traditional experimental paradigms and statistical methods are not well adapted to the analysis of agricultural big data (Coble et al., 2018). Fishers's statistical methods and associated experimental designs are predicated on taking a small sample from a large population, whereas big data will often include very large samples and might at times include the entire population. Furthermore, big data is often associated with high levels of noise, heterogeneity, spurious correlations and incidental endogeneity.

Machine learning presents alternative options for the analysis of big data (Coble et al., 2018). Such algorithms mimic human intelligence by first learning to recognise structures and patterns within sometimes complex datasets and then to use the acquired model, which is akin to human experience, to make predictions about future events. A major advantage of machine learning algorithms for the analysis of big data is that they do not rely on applying user specified models to the data, but instead discern their own rules for the system being scrutinised.

The analysis of agricultural big data using machine learning is becoming increasingly common and recent examples include the prediction of crop type from satellite data, crop yields, irrigation requirements, pest and disease attacks, and weed identification (Pantazi et al., 2016; Kussul et al., 2017; Kamilaris and Preafeta-Boldu, 2018).

Commonly used machine learning tools include Bayesian networks and artificial neural networks (ANNs). A Bayesian network is a machine learning tool that utilises a directed acyclic graph and probability distributions to define and quantify the stochastic dependencies between variables (Pearl, 1988; Koller and Friedman, 2009). Commonly, Bayesian networks can be used to learn a model which describes a complex system. The derived model can act as a substitute for expert human knowledge, and can be used to infer the value of an unknown variable from a given set of known variables (Friedman and Koller, 2003).

In contrast, ANNS are inspired by the physiology of the brain (Haykin, 2007), and use a network of interconnected artificial *in silico* neurones that learn to recognize patterns and relationships among input data, and then use the resulting data model to predict outcomes from new and previously unprocessed input data.

The oil palm industry has embraced the big data paradigm for many years, with estates routinely measuring an enormous array of environmental, agronomic and ecophysiological parameters (Oberthür et al., 2015). The data bank stored by estates presents a valuable yet largely untapped resource to support the development of sustainable palm oil management strategies.

Table 1
Summary of parameters used in the three Bayesian networks.

Parameter name	Description
FFB	Fresh fruit yield in year of harvest ($t{\cdot}ha^{-1}$)
FFB.1	Fresh fruit yield in the year prior to the year of harvest ($t{\cdot}ha^{-1}$)
FFB.2	Fresh fruit yield in the year two years prior to the harvest ($t{\cdot}ha^{-1}$)
ABW	Average weight of fruit bunches in the year of harvest (kg)
ABW.1	Average weight of fruit bunches in the year prior to the year of harvest (kg)
ABW.2	Average weight of fruit bunches in the year two years prior to the harvest (kg)
ESTATE	Identity of the estate from which the data is collected
RAINFALL	Total rainfall in the year of harvest (mm)
RAINFALL.1	Total rainfall in the year prior to the year of harvest (mm)
RAINFALL.2	Total rainfall two years prior to the year of harvest (mm)
SUM_NPKMg_IN	Total fertiliser application in the year of harvest ($kg{\cdot}ha^{-1}$)
SUM_NPKMg_IN.1	Total fertiliser application in the year prior to the year of harvest ($kg{\cdot}ha^{-1}$)
SUM_NPKMg_IN.2	Total fertiliser application in the year two years prior to the harvest ($kg{\cdot}ha^{-1}$)
SMG	Soil management group: classes A, B, C, D and F
TREEAGE	Age of tree in the year of harvest (years)
BUNCH_HA	Density of bunches in the year of harvest ($bunches{\cdot}ha^{-1}$)
BUNCH_HA.1	Density of bunches in the year prior to the year of harvest ($bunches{\cdot}ha^{-1}$)
BUNCH_HA.2	Density of bunches in the year two years prior to the harvest ($bunches{\cdot}ha^{-1}$)
N.17.1	Mean foliar nitrogen content in the 17th frond in the year prior to harvest (% dry matter)
K.17.1	Mean foliar potassium content in the 17th frond in the year prior to harvest (% dry matter)
P.17.1	Mean foliar phosphorous content in the 17th frond in the year prior to harvest (% dry matter)

Oil palm research has developed various machine learning tools to assist the industry including, for example, genomic selection on plant breeding programs (Kwong et al., 2017), the identification of yield recording errors (Pushparani et al., 2018), and fruit ripeness (Bensaeed et al., 2014). Despite these applications of machine learning, and despite the availability of plantation level big data resources, the potential for machine learning resources to predict oil palm yields from commercial big data collections has yet to be explored.

Bayesian networks have great potential for the analysis of big data collected from commercial oil palm estates because: (1) Bayesian networks can integrate both categorical and continuous data, so optimising the full data set (Scutari, 2010); (2) the constructed network shows dependencies between parameters that both validate the learnt network by cross-referencing with pre-existing expert knowledge, or construct new hypothesis through the detection of undiscovered relationships between parameters; (3) Bayesian networks can handle incomplete datasets efficiently (Bressan et al., 2009) and most significantly; (4) the output from a Bayesian network is the level of probability or "belief" that an outcome will occur; managers could easily comprehend and implement probability framed predictions into their estate management. The probability orientated output from Bayesian networks is feature that may be particularly important to support learning processes of estate managers (Tenenbaum, 1999).

In this study, we explore how Bayesian networks can be trained from data sets collected through routine management from commercial oil palm estates, and compared their performance against results from ANNs trained on the same data. A subsequent proof-concept-study predicted yield functions from a range of simple hypothetical situations to demonstrate how trained Bayesian networks could assist estate managers formally represent expectations of future estate productivity under contrasting scenarios.

Table 2
Summary of the five soil (SMG) classes occurring across the estates.

SMG class	Soil depth	Fertility status	Drainage, flooding and moisture status	Soil texture	Dominant slope classes	Comment
A	Moderate to deep	Low to moderate	Well drained, moisture stress	Sandy clay	2–3	Phosphorus fixing
B	Deep	Low to moderate	Well drained	Sandy clay/sandy clay loam	1–2	Soil erosion risk
C	Deep	Low to moderate	Flooding and imperfect to poorly drained	Sandy clay/sandy clay loam		
D	Shallow to moderate	Very low	Prone to flooding and moisture stress	Sandy soils with cement layer	2–3	Cement layer with poor rooting and anchorage
F	Shallow to moderate	Low	Very poorly drained	Organic soils		

Table 3
The number of blocks retained for data processing after filtering for tree age and poor quality data as absolute counts and proportions of the total block numbers.

	2011	2012	2013	2014	2015
Blocks available after filtering poor quality data	68	256	293	272	379
Proportion (%) of total block number	15.26	57.2	65.5	60.8	84.7

Table 4
Annual rainfall data (mm) for each estates (2009 to 2015, NA – no data were available).

	2009	2010	2011	2012	2013	2014	2015
Estate 1	2400	4188	2169	2482	2989	2687	2096
Estate 2	NA	3203	2297	2132	2388	1882	1380
Estate 3	NA	3669	1912	2324	2467	2370	2073

Table 5
Mean and standard deviation of total annual fertiliser applications ($kg{\cdot}ha^{-1}$) for each estate.

	Estate 1	Estate 2	Estate 3
Mean	498	489	467
Standard deviation	108	64	56

Table 6
Summary of the number of blocks of each tree age used in each Bayesian network.

Tree age (years)	6	7	8	9	10	11	12	13	14	15
Number of blocks	54	143	197	247	228	212	112	46	27	2

2. Materials and methods

2.1. Data resources

Data for this study were collected from three commercial estates located in Kalimantan, Indonesia. The study area included some 447 individual blocks with a total area planted to palm of about 19809 ha.

Table 8
Summary of the representation of five SMG classes across the three estates.

SMG class	A	B	C	D	F
Estate 1	46	116	171	83	8
Estate 2	24	154	309	0	0
Estate 3	4	76	242	30	3

Both spatial (topography maps in the form of digital elevation models, soil maps) and non-spatial data (basic site information, daily rainfall and soil survey data) were provided by the estates at the start of the study. The soil survey of the estates was conducted in June 2007. Rainfall data were collected for the period 2004–2015. The fresh fruit bunch (FFB) yield and management data were collected from 2007 till 2015.

Data was manually filtered post collection to remove poor quality observations (Cook et al., 2014) and to restrict the dataset to mature trees only.

2.2. Bayesian network learning

The logic and algorithms underlying the construction of Bayesian networks have been well discussed elsewhere (Bressan et al., 2009; Nagarajan et al., 2013; and Pearl, 1988), but will be reviewed briefly here. The first stage in learning a Bayesian network is to deploy an algorithm which learns the structure of the network by inferring causal relationships between variables within a dataset, and to plot these relationships in a directed acyclic graph comprising nodes connected by arcs. Within each inferred relationship, the independent variable is termed the *parent* node, while the dependent variable is termed the *child* node. The second stage in learning a Bayesian network is to fit parameters to the arcs. The value of the parent node is used to compute the probability density for the child node.

Once the network has been constructed, a subset of nodes with known values can be used to infer the value of a single unobserved target node. The parents of the target node affect its value through their direct conditional dependency. The child nodes are conditionally dependent upon the target node. This dependency can be reversed to infer a value to the target. The child nodes can be affected by one or more alternative parents, so the value of these alternative parents must also be considered when inferring the value of the target from a child node.

Table 7
Summary of the mean, maximum, minimum and standard deviation for the continuous variables derived from blocks used in the three networks (FFB fresh fruit bunches, ABW average bunch weight, BUNCH_HA bunches per hectare, foliar N, foliar P, foliar K).

	FFB ($t{\cdot}ha^{-1}$)	ABW (kg)	BUNCH_HA ($bunches{\cdot}ha^{-1}$)	Foliar N content (%) in frond 17	Foliar P content (%) in frond 17	Foliar K content (%) in frond 17
Mean	25.1	14.7	1791	2.6	0.158	0.936
Max	44.1	23.6	5263	3.2	0.262	1.46
Min	8.0	4.9	509	1.9	0.106	0.16
SD	4.6	3.7	464	0.23	0.016	0.159

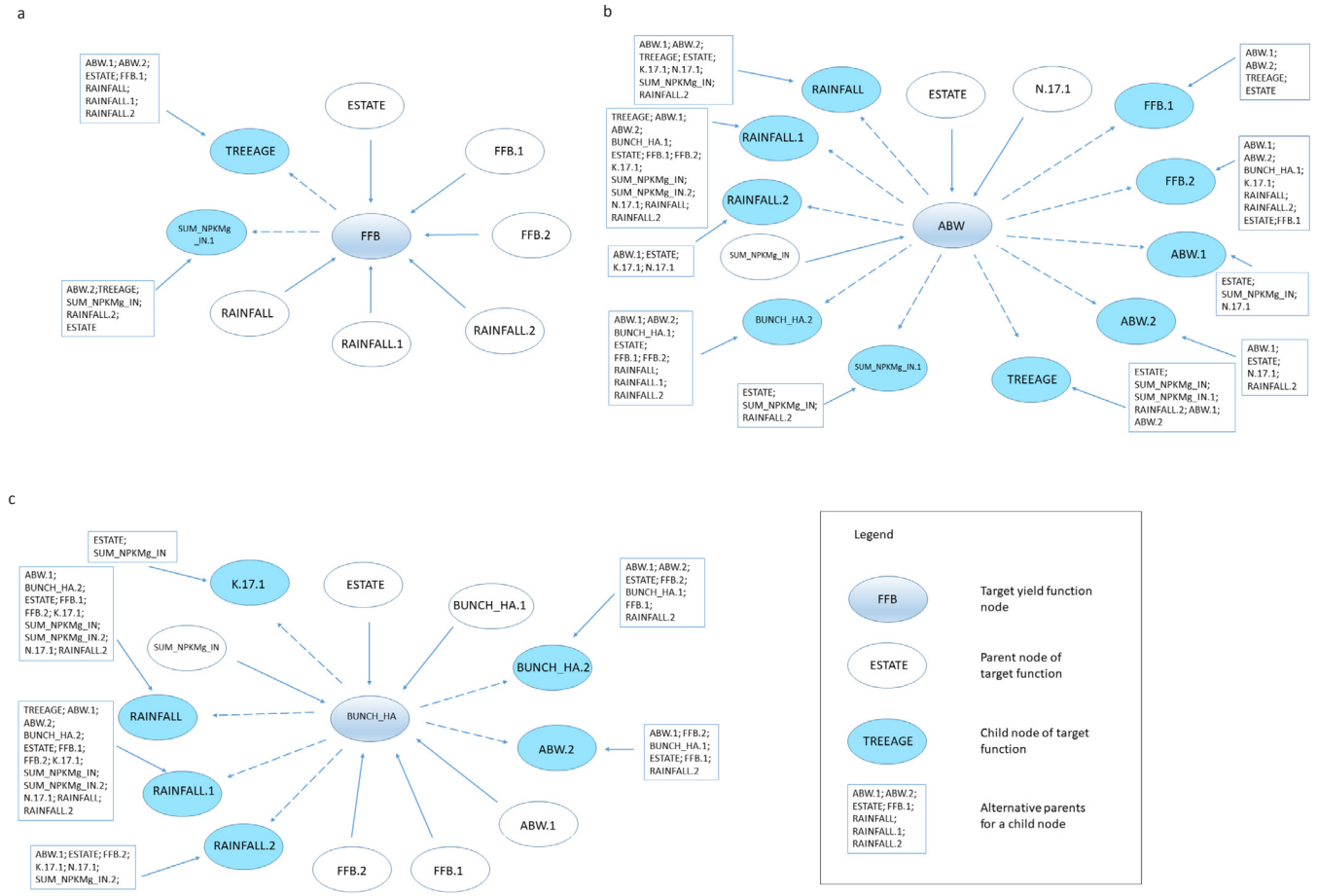

Fig. 1. The Markov blankets for (a) the FFB; (b) the ABW; and (c) the BUNCH_HA nodes extracted from the network trained for each of these yield functions (FFB fresh fruit bunches, ABW average bunch weight, BUNCH_HA bunches per hectare).

Table 9
The mean (standard deviation) frequency with which the Bayesian networks assigned a block to the correct yield class using threshold values of mean, 25th percentile, and 75th percentile of FFB, ABW and BUNCH_HA within each test dataset.

	25th percentile yield threshold	Mean yield threshold	75th percentile yield threshold
FFB network	85.6 (3.7)	78.3 (3.7)	79.0 (3.4)
ABW network	94.9 (3.2)	94.9 (2.2)	92.4 (1.9)
BUNCH_HA network	89.1 (2.3)	85.6 (2.8)	89.0 (2.2)

Table 10
Comparison of the mean r-squared (standard deviation) from Bayesian network and ANNs for 3 yield functions.

Yield function	Bayesian network	ANN
FFB	0.6 (0.051)	0.6 (0.141)
ABW	0.9 (0.011)	0.9 (0.022)
BUNCH_HA	0.8 (0.027)	0.5 (0.269)

The suite of nodes that influence the target node (the parents, children and alternative parents of the children) are termed the Markov Blanket for the target node, and the Markov Blanket represents the sum of knowledge required to predict the value of that target node. The inclusion of multiple and parallel influences on the target node typically allow for accurate predictions even from noisy data.

A Bayesian network which represented the conditional dependencies between the variables in each dataset was constructed using the *bnlearn* package constructed for the R software environment (Scutari, 2010). As learning a Bayesian network is computationally complex, or NP-hard, the heuristic "hill-climbing" algorithm was used to learn the network structure by efficiently searching for an approximate rather than perfect solution (Gamez et al., 2010; Scutari, 2017). The network parameters were computed using the maximum likelihoods method as this method can parameterize both continuous and discrete data (Nagarajan et al., 2013).

The objective of the networks was to predict the future performance of three yield functions; total yield of fresh fruit bunches in each block (FFB), the average weight of harvested bunches (ABW), and the total bunch number per hectare (BUNCH_HA) based on past and current management and environmental records along with past records of yield functions. The networks therefore included management and environmental data from across a three year window, including the year of harvest year and the two years immediately prior. Restricting the window to three years suppressed the influence of very young trees with unusually low productivity from the networks.

The range of parameters used across three networks are described in Table 1. As the three key yield functions (FFB, ABW, BUNCH_HA) are recorded simultaneously at harvest, it is impossible to know one yield function value in advance of any other. It is therefore meaningless to utilise any yield function variables from the year of harvest to make predictions regarding other yield functions. Thus, all yield function data from the year of harvest other than the target function were excluded from each network. For example, a network trained to predict FFB excluded measurements of ABW and BUNCH_HA from the harvest year. The resulting networks were deployed to predictions for the FFB,

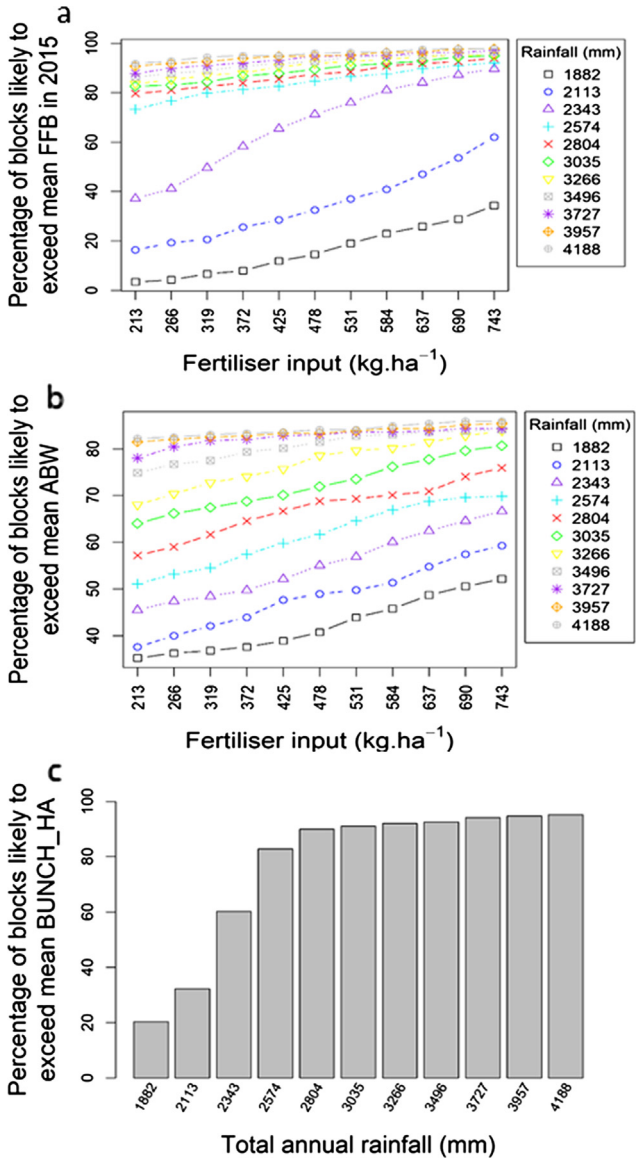

Fig. 2. The percentage of blocks predicted to exceed (a) mean FFB yield (t·ha^{-1}); (b) mean ABW weight (kg); and (c) BUNCH_HA (bunches·ha^{-1}) in 2015 following hypothetical rainfall levels, and also fertiliser inputs and for FFB and ABW variables in the previous year.

BUNCH_HA, and ABW yield functions for a given set of environmental, management and agronomic parameters.

The accuracy of each network was tested using ten-fold cross-validation (10-FOLD CV) (Lever et al., 2016). The 10-FOLD CV method involves repeatedly dividing the dataset into two subsets, the first subset is used to train a network. The second dataset is then fed into the network to make predictions of yield based on the values of all other variables. The accuracy of the network was described by comparing predicted yield functions with observations within the test dataset. In this study, the 10-FOLD CV was performed ten times, on each occasion the test dataset was composed of a unique random selection of 10% of the total data. The accuracy of the Bayesian networks was assessed by (1) determining the precision with which the network predicted that yield functions would exceed three thresholds (25%, mean, or 75% of the observed yield function values) and; (2) comparing the most probable yield value (*argmax*) against actual yield values using linear regression.

2.3. Comparison with an artificial neural network

The Bayesian Network's learning efficiency was compared with results generated from an ANN. The ANNs were created using the *neuralnet* R package (Günther and Fritsch, 2010). ANNs were trained for each yield function, with data transformed so that each SMG was represented by an integer (A = 1, B = 2, C = 3, D = 4, F = 5), and all variables re-scaled to a range between 0 and 1. The ANNs were of a multilayer perceptron architecture, with 2 hidden layers of eight and three nodes, and one single output variable. ANN learning used the resilient backpropagation algorithm which gives robust but rapid learning outcomes (Riedmiller and Braun, 1993; Schmidhuber, 2015). Output from every node was a sigma function, except for the final node which produced a linear output.

2.4. Application

We demonstrated utility of Bayesian networks to support management decisions by computing expectation that each block would achieve a specified yield given prior rainfall and fertilizer inputs. The targets selected were equivalent to the mean of all values observed for 2015 (27.16 t·ha^{-1} FFB; 16.93 kg ABW; 1639 bunches·ha^{-1}), given a range of hypothetical conditions. For each network, we selected different rainfall levels in the preceding year (RAINFALL.1). Fertiliser from the year prior to harvest (SUM_NPKMg_IN.1) was also included for those networks where it was included in the Markov Blanket for the yield function. The selected conditions spanned all observed rainfall and fertiliser inputs, set at intervals of 10% of the total observed range. All other conditions were as recorded for each block during 2015. The outcome was summarised as the proportion of blocks across that are predicted to exceed the target threshold for each set of conditions.

3. Results

3.1. Data description

The experimental area included 5 distinct soil management groups (SMGs) representing a range of soil depth, fertility, textural classes, drainage and moisture status, and topographies (Table 2).

The numbers of blocks retained after filtering for immature trees and poor quality data generally increased over the period used for this study (Table 3), mainly because the number of blocks of sufficient maturity increased over time. The low number of blocks filtered out during 2015 reflects the generally high quality of data collected from mature, productive blocks.

Total annual rainfall (mm) data over the years for which data was used for this study is summarised in Table 4. Rainfall data from each block included historical data from the two previous years, meaning that rainfall data was collected over a longer time frame than the block data (Table 3).

The median annual fertiliser applications made to each estate were very similar, although estate 1 had a slightly higher diversity of application rates (Table 5).

The age of trees used to learn and develop the networks spanned from 6 to 15 years (Table 6).

Crop yield and nutrition variables were highly variable from across the data used in this study (Table 7), with FFB and ABW showing an approximate 5 fold differences between the highest and lowest observed values, and BUNCH_HA showing a 10-fold difference.

There were five different SMG classes; one of these (C) were more common in estate 2 and estate 3, whereas classes A, D and especially F were more common in estate 1 (Table 8).

3.2. Learning Bayesian networks

The Bayesian networks constructed for the FFB, ABW, and

BUNCH_HA data are presented in Fig. 1. The Markov blanket associated with the FFB node included 11 other variables. Six variables were parents of the FFB node (ESTATE, FFB.1, FFB.2, RAINFALL, RAINFALL.1, RAINFALL.2), indicating a direct relationship with FFB variable; and two nodes were included as children (SUM_NPKMg_IN.1 and TREEAGE), indicating that their influence will be modulated by their alternative parents.

Comparing the predicted and actual classification of yield against 3 thresholds showed that the FFB network was highly accurate, assigning predicted FFB to the correct class in over 75% with all target threshold levels (Table 9). The most probable *argmax* yield for FFB showed a good correlation with observed yields (Table 10), indicating that the main drivers of FFB had been well captured.

The Bayesian network for the ABW data (Fig. 1) was more complex, including a Markov blanket of 16 nodes. The ABW node had three parent nodes (ESTATE, SUM_NPKMg_IN, and N.17.1) and ten child nodes (TREEAGE, BUNCH_HA.2, ABW.1, ABW.2, FFB.1, FFB.2, RAINFALL, RAINFALL.1, RAINFALL.2 and SUM_NPKMg_IN.1). The ABW network was very highly accurate, assigning predicted ABW to the correct class in approximately 90% of instances (Table 9), and the ABW *argmax* was very highly correlated with observed data (Table 10), again indicating a very high level of precision.

The network for the BUNCH_HA data (Fig. 1) consisted of 6 parents for the BUNCH_HA node (ESTATE, BUNCH_HA.1, ABW.1, FFB.1, FFB.2, and SUM_NPKMg_IN) along with 6 child nodes (RAINFALL, RAINFALL.1, RAINFALL.2, ABW.2, BUNCH_HA.2, K.17.1). The network predicted yield class accurately in close to 90% of instances (Table 9) and showed a strong correlation between *argmax* and observed data (Table 10).

Comparing the accuracy of Bayesian networks and ANNs shows that both methods gave extremely high precision for the ABW function (Table 10). ANN gave similar but somewhat reduced precision for FFB. The Bayesian network also returned very good precision for BUNCH_HA, but the ANN's output for that function was the least accurate of all predictions

The various parameters that quantified the relationships between the FFB, BUNCH_HA and ABW nodes on the three networks are summarised in (Appendix Tables A1–A3).

3.3. Bayesian network application

The utility of Bayesian networks in supporting management decisions is demonstrated by computing the probability that each block will exceed a target FFB, ABW or BUNCH_HA given a range of hypothetical conditions in the year preceding harvest. We predicted FFB and ABW from RAINFALL.1 and SUM_NPKMg_IN.1. However, as the BUNCH_HA network did not include SUM_NPKMg_IN.1, predictions were derived solely from hypothetical RAINFALL.1 values (Fig. 2).

All three variables showed a clear positive response to increasing RAINFALL.1, and FFB and ABW showed the greatest response at the lowest fertiliser rates. However, historic rainfall inputs brought diminishing returns, with little benefit observed above 2804 mm (FFB and BUNCH_HA) or above 3496 mm (ABW), where most blocks already showed a high probability of exceeding yield targets.

Some blocks showed a low probability of exceeding the target threshold at even the highest inputs in the preceding year, including eight blocks in the FFB, 110 in the ABW, and 17 in the BUNCH_HA networks. Similarly, other blocks showed a high likelihood of exceeding the target threshold at even the lowest historic input levels, including 2 blocks in the FFB network, 164 blocks in the ABW network, and 77 blocks in the BUNCH_HA network. Collectively, these results indicate that, whilst production from most blocks show substantial response to historic rainfall and/or fertiliser inputs, some blocks are insensitive to such variation.

4. Discussion

4.1. Bayesian networks and management support

As external pressures on the oil palm industry increase, the future sustainability of the industry depends on the ability of managers to make informed adaptations to management process. Vital to the support of process change will be the provision of decision support tools to free manager's judgement from intuition based errors and biases.

The analysis performed in this study demonstrate that Bayesian networks can be used to successfully predict yield functions from commercial oil palm estates using data collected as routine estate management practice. The networks created will compliment and extend the predictions made possible with PALSIM simulation model (Hoffmann et al., 2014) in that they utilise a greater diversity of input data, including current and historic management factors, soil type and past rainfall. As such, the Bayesian networks will better predict the impact of climate as well as the effect of differences in management factors and soil conditions, both within and between estates, on production. Furthermore, the Bayesian networks will have two major advantages over the development of novel modelling approaches. First, Bayesian networks utilise pre-existing data, and so do not require any costly field experimentation to compute relationships between input parameters and yield (Hoffmann et al., 2014). Second, they are based on empirical observations obtained directly from the sites of interest, so avoid the problems with extrapolating observations across different geographic locations (Hoffmann et al., 2014). The information generated by the networks can guide a manager's expectations of yield outcomes from across a range of contrasting environmental and agronomic conditions. This information is complementary to benchmarking with PALMSIM-generated estimate of annual ceiling yields, a tool which describes the potential yield ceiling based solely on sunlight and water (Hoffmann et al., 2015), by providing managers with insight to the agronomic or ecophysiological factors underlying any gap between actual potential yield in any year.

All networks predicted yield with high levels of accuracy with both the yield threshold and *argmax's* most probable yield value, indicating that the networks capture the key relations between input variables and yield functions. The ABW and BUNCH_HA networks both returned a higher level of precision than the FFB. The reduced accuracy of the FFB may be due in part to errors in data recording, which is performed at the receiving mill, and is subjected to numerous sources of inaccuracy, including variable efficiencies in crop harvesting; weight loss between harvest and mill-processing; and human or technical errors at the point of weighing. The development of techniques to reduce the impact of data inaccuracies within the FFB data set presents a challenge for future research. A likely first step will be to identify and remove poor quality FFB data points during training and testing.

Comparing the precision of the two learners reveals that the Bayesian networks compare well against the ANNs, with the Bayesian methods equalling or exceeding the ANN's accuracies for all measures. The reason for Bayesian network's superior performance in this study is not clear, but similar findings have been reported elsewhere (e.g. Correa et al., 2009).

4.2. Relationships between factors within the Bayesian networks

Examining the parameters included in the FFB, ABW and BUNCH_HA networks gives an indication of the factors underlying these yield functions. All networks included rainfall from the current and past years, indicating that annual rainfall influences production in both

current season and subsequent years (Cock et al., 2016). All networks included historic fresh fruit yields and bunch weight (FFB.1, FFB.2, ABW.1 and ABW.2), demonstrating that past productivity is an indicator of future yield. SUM_NPKMg_IN was also present within each network indicating that fertiliser applied during the year of harvest impacted yield functions. The inclusion of the TREEAGE variable in all networks demonstrates that tree maturity also affects yield (Corley and Tinker, 2015; and Mahamooth et al., 2011). The three estates incorporated into this study can be distinguished by a range of geographic and environmental factors, many of which are described by other variables in the networks, including annual rainfall levels and soil SMG class. However, the inclusion of the ESTATE variable indicates that other unknown and possibly management related estate-associated factors influence productivity; such factors are worthy of future research.

The inclusion of SUM_NPKMg_IN.1 as a child of the target node in the FFB and ABW networks indicates that the influence of fertiliser applied in the year prior to harvest is modulated by other factors, including the level of fertiliser applied in the year of harvest and rainfall levels two years earlier.

The critical function of TREEAGE in determining yield is demonstrated by its role as a child node of both FFB and ABW yield functions. Furthermore, its observed role as an alternative parent on SUM_NPKMg_IN.1 (FFB network) demonstrates that tree age impacts the fertiliser response. Similarly, the historic bunch weight (ABW.2) also moderates historical fertiliser effects (SUM_NPKMg_IN.1) in the same network; this apparent relationship between historic variables would benefit from further scrutiny.

The inclusion of historic fertiliser variables in each network demonstrates that prior conditions influence yield functions; similar relationships have been previously reported by Goh and Härdter (2003). The ABW and BUNCH_HA networks also included two leaf foliar nutrition measurements from the year prior to harvest (N.17.1 and K.17.1), confirming again the role of past crop nutrition on current yield.

The networks demonstrate that past bunch numbers affect the ABW and BUNCH_HA functions, but the negative relationship between BUNCH_HA.2 and ABW reiterates the previously reported inverse relationship between these parameters (Corley and Tinker, 2015).

4.3. Applications of the Bayesian networks to hypothetical rainfall and fertiliser conditions

The utility of the Bayesian networks was demonstrated by predicting the probabilities that blocks would exceed target FFB, ABW or BUNCH_HA thresholds for a given range of fertiliser and rainfall inputs during the years preceding the harvest.

4.3.1. Rainfall conditions

All yield functions showed a clear positive response to rainfall in the year prior to harvest. However, increasing rainfall brought a diminishing response from each function. For example, the BUNCH_HA's response approached a maximum at 2804 mm where approximately 90% of blocks were already exceeding the threshold. Similarly, more than of 2574 mm brought little benefits to the FFB function, and the ABW's response diminished above 3496 mm. Similar relationships have been discussed previously by Corley and Tinker (2015).

4.3.2. Fertiliser inputs

Both FFB and ABW yield functions responded positively to fertiliser applications in the year prior to harvest. The fertiliser's predicted positive response at low rainfall suggests that nutrient management may be used to offset the negative impacts of water deficits.

The diminishing response to fertiliser at high rainfall indicates that most blocks will give above average yield following wet years, even with low prior fertiliser. We do not have an explanation for this, although possible explanations include biological stimulation rising from high growth in previous wet years, or low yield taken in the previous years due to flood related crop losses.

The response to fertiliser from both the FFB and ABW functions declined with increasing rainfall with an asymptote being approached at 2574 mm for the FFB network, and 3496 mm for the ABW network. At these rainfall levels, most blocks are likely to exceed the target yield at the lowest SUM_NPKMg_IN.1 levels, so few respond to additional fertiliser.

4.3.3. Management implications

The yield function's predicted responses can be assigned to three main classes, each has a unique implication for management. Some blocks show a high probability of exceeding the mean FFB, ABW or BUNCH_HA function, even at the lowest RAINFALL.1 or SUM_NPKMg_IN.1. Management will therefore have limited potential to change the productivity of these high yielding blocks, and they will provide predictably high yield under a diverse range of conditions. In contrast, a different set of blocks showed a low probability of exceeding the thresholds at even the highest RAINFALL.1 and SUM_NPKMg_IN.1 levels; these blocks will therefore be consistently low yielding. These blocks will be highly resistant to improvement through management unless the manager can identify impediments for such low yields. For the remaining blocks, the probability of exceeding the yield threshold responded to RAINFALL.1 levels and, for the FFB and ABW networks, the SUM_NPKMg_IN.1 levels. Focussing management resources on these blocks will bring the greatest production responses, especially when matching SUM_NPKMg_IN.1 with RAINFALL.1.

5. Conclusions

This study has demonstrated that Bayesian networks based on data collected during routine management of oil palm plantations can be successfully trained to predict yield functions with a high degree of precision. The resulting networks can be deployed to provide robust predictions regarding the probability of achieving a specified yield threshold following a given set of environmental parameters and management strategies. Together with generalized predicted tools such as PALMSIM, such machine-learning methods will provide a vital resource in aiding plantation managers to make rational and evidence-based decisions when formulating strategic and process orientated changes to management in response to emerging social, political and environmental pressures within the oil palm industry. Furthermore, the key finding that machine learning can extract value from complex datasets will have broad potential for the fast emerging field of big data in broader agricultural industries.

Acknowledgements

Results reported here originated from an ongoing collaborative project between IPNI Southeast Asia Program and Wilmar International Limited, with funding provided by IPNI and Canpotex Limited. We would like to thank the two anonymous referees for their helpful comments on the manuscript.

Appendix

Tables A1–A3.

Table A1
Summary of coefficients from the FFB network for (a) parent nodes for the FFB variable; (b) parents of the SUM_NPKMg_IN.1 node and; (c) parents of the TREEAGE node.

ESTATE	ESTATE 1	ESTATE 2	ESTATE 3
A1.a			
(Intercept)	33.8	−6.55	−15.4
RAINFALL	−8.93E−03	−6.28E−03	−6.32E−03
RAINFALL.1	−3.47E−04	9.90E−03	1.50E−02
RAINFALL.2	−2.87E−04	1.71E−03	4.02E−03
FFB.1	31.0	53.5	14.0
FFB.2	32.4	18.7	34.1
A1.b			
(Intercept)	574.3	745.5	367.3
TREEAGE	−28.2	6.925	5.652
SUM_NPKMg_IN	−0.033	−0.292	−0.136
RAINFALL.2	−0.004	−0.055	0.005
FFB	3.008	2.403	3.147
ABW.2	12.872	−7.347	2.686
A1.c			
(Intercept)	17.703	7.189	0.51
RAINFALL	−0.002	−6.0−E04	−4.95E−04
RAINFALL.1	−0.001	−1.0E−04	1.86E−03
RAINFALL.2	−0.001	−3.0E−04	2.48E−04
FFB	0.0432	0.004	−4.57E−04
FFB.1	−0.058	−0.031	−6.07E−03
ABW.1	0.156	0.158	0.17
ABW.2	0.225	0.289	0.226

Table A2
Summary of coefficients from the ABW network for (a) parent nodes for the ABW variable; (b) parents of the FFB.1 child; (c) parents for the FFB.2 child node; (d) parents of the ABW.1 node; (e) parents of the ABW.2 node; (f) parents of the TREEAGE; (g) parents of the SUM_NPKMg_IN.1 node; (h) parents of the BUNCH_HA.2 node; (i) parents of the RAINFALL node; (j) parents of the RAINFALL.1 node; (k) parents of the RAINFALL.2 node.

ESTATE	ESTATE 1	ESTATE 2	ESTATE 3
A2.a			
(Intercept)	33.9	35.0	36.4
N.17.1	−8.858	−5.728	−9.017
SUM_NPKMg_IN	3.3E−03	−5.8E−03	5.0E−3
A2.b			
(Intercept)	18.6	33.4	25.9
TREEAGE	−1.405	−1.485	−0.650
ABW	0.066	−0.662	−1.481
ABW.1	1.409	0.431	1.580
ABW.2	−0.057	0.916	0.60
A2.c			
Intercept	−10.9	16.2	−6.753
RAINFALL	5.30E−03	2.77E−03	1.19E−02
RAINFALL.2	4.33E−04	1.02E−03	3.77E−03
FFB.1	8.01E−01	2.55E−01	1.39E+00
ABW	1.01	0.466	1.49
ABW.1	−1.96	−0.133	−4.13
ABW.2	1.38	−0.324	1.57
K.17.1	−2.05	−4.30	−0.0012
BUNCH_HA.1	−1.47E−03	−1.12E−03	−1.30E−02
A2.d			
Intercept	−4.089	4.923	8.698
SUM_NPKMg_IN	2.38E−04	−0.005	−0.008
ABW	0.892	0.94	0.902
N.17.1	1.396	−1.027	−1.93

(*continued on next page*)

Table A2 (*continued*)

ESTATE	ESTATE 1	ESTATE 2	ESTATE 3
A2.e			
(Intercept)	0.31	5.535	4.77
RAINFALL.2	−8.2E−04	−0.001	−6.4E−04
ABW	0.076	0.258	0.073
ABW.1	0.766403	0.6925	0.657275
N.17.1	0.79569	−1.509	−0.653
A2.f			
(Intercept)	7.023	5.461	4.436
SUM_NPKMg_IN	−0.003	−0.001	−0.001
SUM_NPKMg_IN.1	−0.003	5.72E−04	−2.2−E04
RAINFALL.2	−1.3E−04	−5.2E−04	−2.0−E04
ABW	0.154	−0.069	0.111
ABW.1	0.122	0.263	0.209
ABW.2	0.181	0.276	0.121
A2.g			
(Intercept)	443.9	776.1	411.3
SUM_NPKMg_IN	0.066	−0.295	−0.199
RAINFALL.2	−0.001	−0.041	0.013
ABW	1.830	−1.873	8.648
A2.h			
(Intercept)	1.48E+03	3.45E+02	1.97E+03
RAINFALL	−0.222	−4.96E−02	3.19E−04
RAINFALL.1	−8.74E−02	3.67E−02	−0.354
RAINFALL.2	−5.46E−02	8.65E−02	−8.01E−02
FFB.1	−59.0	−48.3	−49.9
FFB.2	117.0	67.6	84.9
ABW	−2.96	−11.8	−23.0
ABW.1	122.2	90.2	135.6
ABW.2	−225.5	−122.8	−199.7
BUNCH_HA.1	0.688	0.822	0.639
A2.i			
(Intercept)	770.6	906.5	1972.0
TREEAGE	17.355	−165.7	−52.8
SUM_NPKMg_IN	0.185	1.519	1.139
RAINFALL.2	−0.10	0.141	−0.047
ABW	−107.4	95.03	28.5
ABW.1	−5.074	−100.6	−18.894
ABW.2	134.3	61.99	−17.979
N.17.1	718.7	460.0	155.7
K.17.1	−35.6	−401.0	−28.3
A2.j			
(Intercept)	8617.5	1639.4	3611.79
TREEAGE	−264.0	7.249	19.078
SUM_NPKMg_IN	0.727	0.038	0.188
SUM_NPKMg_IN.2	0.459	−0.019	−0.159
RAINFALL	−1.495	0.282	−0.282
RAINFALL.2	−0.716	−0.059	−0.282
FFB.1	–23.2	–33.7	6.724
FFB.2	14.091	19.995	−1.179
ABW	44.7	−12.554	22.7
ABW.1	17.42	101.1	15.3
ABW.2	17.3	−70.2	−63.0
N.17.1	413.5	−100.6	50.8
K.17.1	−316.1	−110.1	15.4
BUNCH_HA.1	0.066	0.214	−0.130
A2.k			
(Intercept)	5323.7	1221.9	3547.6
ABW	−193.5	−177.0	−326.4
ABW.1	218.8	113.6	278.4
N.17.1	−204.8	811.6	275.6
K.17.1	−2171.1	427.8	−724.0

Table A3
Summary of coefficients from the BUNCH_HA network for (a) parent nodes for the BUNCH_HA variable; (b) parents of the BUNCH_HA.2 child node; (c) parents for the ABW.2 child node; (d) parents of the K.17.1 child node; (e) parents of the RAINFALL child node; (f) parents of theRAINFALL.1 child node ; (g) parents of the RAINFALL.2 child node.

ESTATE	ESTATE 1	ESTATE 2	ESTATE 3
A3.a			
(Intercept)	82.12	378.60	2035.78
SUM_NPKMg_IN	0.02	−0.93	−1.92
FFB.1	−65.01	−32.90	−7.86
FFB.2	−7.64	18.19	9.72
ABW.1	92.83	32.51	−17.02
BUNCH_HA.1	1.15	0.88	0.44
A3.b			
(Intercept)	397.1	197.7	555.4
RAINFALL.2	0.010	0.082	0.035
FFB.1	−63.7	−50.3	−51.0
FFB.2	113.5	65.8	82.8
ABW.1	145.4	86.4	103.7
ABW.2	−235.4	−122.8	−177.0
BUNCH_HA	−0.034	0.069	0.074
BUNCH_HA.1	0.755	0.818	0.688
A3.c			
(Intercept)	2.620	−2.494	−0.092
RAINFALL.2	−6.0E−04	−0.001	−6.7.E−04
FFB.1	−0.073	−0.150	−0.087
FFB.2	0.085	0.005	0.072
ABW.1	0.790	1.250	0.906
BUNCH_HA	−4.9E−04	−0.002	−7.6E−04
BUNCH_HA.1	3.3E−04	0.004	0.001
A3.d			
(Intercept)	1.05	1.275676279	1.212
SUM_NPKMg_IN	−1.6E−04	−6.1E−04	−6.4E−04
BUNCH_HA	−5.4E−05	−4.7E−05	−1.3E−0.5
A3.e.			
(Intercept)	750.8	1006.5	1820.9
SUM_NPKMg_IN	0.055	1.512	0.803
SUM_NPKMg_IN.2	−0.419	−0.874	−0.019
RAINFALL.2	−0.066	0.197	−0.071
FFB.1	9.168	15.136	−2.739
FFB.2	43.3	33.0	−11.329
ABW.1	−76.7	−78.0	−1.852
N.17.1	975.8	464.6	135.9
K.17.1	199.2	−307.0	−73.7
BUNCH_HA	−0.079	−0.319	−0.144
BUNCH_HA.2	−0.337	−0.162	0.269
A3.f			
(Intercept)	9500.469	1674.4	3605.1
TREEAGE	−246.2	5.196	24.0
SUM_NPKMg_IN	0.705	0.046	0.203
SUM_NPKMg_IN.2	0.448	−0.004	−0.091
RAINFALL	−1.649	0.304	−0.204
RAINFALL.2	−0.748	−0.077	−0.297
FFB.1	−16.120	−23.4	−5.767
FFB.2	37.8	13.190	19.441
ABW.1	30.1	66.2	54.7
ABW.2	−11.423	−51.7	−99.9
N.17.1	438.5	−107.9	57.3
K.17.1	−389.2	−100.4	8.734
BUNCH_HA	0.030	0.143	0.052
BUNCH_HA.2	−0.188	0.059	−0.237
A3.g			
(Intercept)	1085.1	−1048.6	−2004.6
SUM_NPKMg_IN.2	0.317	0.520	1.607
FFB.2	−38.5	1.025	−1.340
ABW.1	135.4	−1.334	73.1
N.17.1	496.5	755.0	449.2
K.17.1	−1749.7	496.6	−313.0
BUNCH_HA	0.571	0.574	1.013

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.compag.2018.06.006.

References

Antle, J.M., Jones, J.W., Rosenzweig, C.E., 2017. Next generation agricultural system data, models and knowledge products: Introduction. Agric. Syst. 155, 186–190. http://dx.doi.org/10.1016/j.agsy.2016.09.003.

Bensaeed, O.M., Shariff, A.M., Mahmud, A.B., Shafri, H., Alfatn, M., 2014. Oil palm fruit grading using a hyperspectral device and machine learning algorithm. IOP Conference Series: Earth and Environmental Science. 20: 012017. doi.ooorg/10.1088/1755-1315/20/1/012017.

Bressan, G.M., Oliveira, V.A., Hruschka, E.R., Nicoletti, M.C., 2009. Using Bayesian networks with rule extraction to infer the risk of weed infestation in a corn-crop. Eng. Appl. Artificial Intelligence 22, 579–592. http://dx.doi.org/10.1016/j.engappai.2009.03.006.

Carlson, K.M., Curran, L.M., Asner, G.P., Pittman, A.M., Trigg, S.N., Adeney, J.M., 2013. Carbon emissions from forest conversion by Kalimantan oil palm plantations. Nat. Climate Change 3, 283. http://dx.doi.org/10.1038/nclimate1702.

Chi, M., Plaza, A., Benediktsson, J.A., Sun, Z., Shen, J., Zhu, Y., 2016. Big data for remote sensing: challenges and opportunities. Proc. IEEE 104, 2207–2219. http://dx.doi.org/10.1109/JPROC.2016.2598228.

Chi, M., Plaza, A., Benediktsson, J.A., Sun, Z., Shen, J., Zhu, Y., 2016. Big data for remote sensing: challenges and opportunities. Proc. IEEE 104, 2207–2219. http://dx.doi.org/10.1109/JPROC.2016.2598228.

Coble, Keith H., Mishra, Ashok K., Ferrell, Shannon, Griffin, Terry, 2018. Big data in agriculture: a challenge for the future. Appl. Econ. Perspect. Policy 40, 79–96. http://dx.doi.org/10.1093/aepp/ppx056.

Cock, J., Kam, S.P., Cook, S., Donough, C., Lim, Y.L., Jines-Leon, A., Lim, C.H., Primananda, S., Yen, B.T., Mohanaraj, S.N., Samosir, Y.M.S., 2016. Learning from commercial crop performance: Oil palm yield response to management under well-defined growing conditions. Agric. Syst. 149, 99–111. http://dx.doi.org/10.1016/j.agsy.2016.09.002.

Cook, S., Lim, C.H., Mohanaraj, S.N., Samosir, Y.M.S., Donough, C., Oberthür, T., Lim, Y.L., Cock, J., Kam, S.P., 2014. Palm oil at the crossroads: the role of plantation intelligence to support change, profit and sustainability. The Planter, Kuala Lumpur 90, 563–575.

Corley, R.H.V., Tinker, P.B., 2015. The Oil Palm, fifth ed. Wiley.

Correa, M., Bielza, C., Pamies-Teixeira, J., 2009. Comparison of Bayesian networks and artificial neural networks for quality detection in a machining process. Expert Syst. Appl. 36, 7270–7279. http://dx.doi.org/10.1016/j.eswa.2008.09.024.

Friedman, N., Koller, D., 2003. Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks. Mach. Learning 50, 95–125. http://dx.doi.org/10.1023/A:102024991.

Gamez, J.A., Mateo, J.L., Puerta, J.M., 2010. Learning Bayesian networks by hill climbing: efficient methods based on progressive restriction of the neighbourhood. Data Min. Knowl. Disc. 22, 106–148. http://dx.doi.org/10.1007/s10618-010-0178-6.

Goh, K.J., Härdter, R., 2003. General oil palm nutrition. In: Fairhurst, T.H., Härdter, R. (Eds.), Oil Palm: Management for Large and Sustainable Yields. Potash & Phosphate Institute, Singapore.

Günther, F., Fritsch, S., 2010. Neuralnet: training of neural networks. The R J. 2, 30–38.

Haykin, S., 2007. Neural Networks: A Comprehensive Foundation, third ed. Prentice-Hall Inc, Upper Saddle River, NJ, USA.

Hoffmann, M.P., Vera, A.C., van Wijk, M.T., Giller, K.E., Oberthür, T., Donough, C., Whitbread, A.M., 2014. Simulating potential growth and yield of oil palm (Elaeis guineensis) with PALMSIM: Model description, evaluation and application. Agric. Syst. 131, 1–10. http://dx.doi.org/10.1016/j.agsy.2014.07.006.

Hoffmann, M.P., Donough, C., Oberthür, T., Vera, A.C., van Wijk, M.T., Lim, C.H., Asmono, D., Samosir, Y., Lubis, A.P., Moses, D.S., Whitbread, A.M., 2015. Benchmarking yield for sustainable intensification of oil palm production in Indonesia using PALMSIM. The Planter, Kuala Lumpur 91, 81–96.

Lever, J., Krzywinski, M., Altman, 2016. Points of significance: model selection and overfitting. Nat. Methods 13, 703–704.

Kamilaris, A., Kartakoullis, A., Prenafeta-Boldú, F.X., 2017. A review on the practice of big data analysis in agriculture. Comput. Electron. Agric. 143, 23–37.

Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. Comput. Electron. Agric. 147, 70–90. http://dx.doi.org/10.1016/j.compag.2017.09.037.

Koller, D., Friedman, N., 2009. Probabilistic Graphical Models: Principles and Techniques. MIT Press.

Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. Remote Sens. Lett. 14, 778–782. http://dx.doi.org/10.1109/LGRS.2017.2681128.

Kwong, Q.B., Teh, C.K., Ong, A.L., Chew, F.T., Mayes, S., Kulaveerasingam, H., Tammi, M., Yeoh, S.H., Appleton, D.R., Harikrishna, J.A., 2017. Evaluation of methods and marker systems in genomic selection of oil palm (Elaeis guineensis Jacq.). BMC Genet. 18, 107. http://dx.doi.org/10.1186/s12863-017-0576-5.

Lokers, R., Knapen, R., Janssen, S., van Randen, Y., Jansen, J., 2018. Analysis of big data technologies for use in agro-environmental science. Environ. Modell. Software 84, 494–504. http://dx.doi.org/10.1016/j.envsoft.2016.07.017.

Mahamooth, T.N., Gan, H.H., Kee, K.K., Goh, K.J., 2011. Water requirements and cycling of oil palm. In: Goh, K.J., Chiu, S.B., Paramananthan, S. (Eds.), Agronomic Principles and Practices of Oil Palm Cultivation. Agricultural Crop Trust, Petaling Jaya, pp. 89–144.

Morota, G., Ventura, R.V., Silva, F.F., Koyama, M., Fernando, S.C., 2018. BIG DATA ANALYTICS AND PRECISION ANIMAL AGRICULTURE SYMPOSIUM: Machine learning and data mining advance predictive big data analysis in precision animal agriculture. J. Anim. Sci. 96, 1540–1550. http://dx.doi.org/10.1093/jas/sky014.

Nagarajan, R., Scutari, M., Lèbre, S., 2013. Bayesian Networks in R with Applications in Systems Biology. Springer, New York.

Oberthür, T., Cook, S., Donough, C., Cock, J., Pheng, S., Li, Y., 2015. Inteligencia de Plantaciones de palma de aceite: análisis de datos de producción para la toma de decisiones agronómicas efectivas y el manejo de fertilizantes. Palmas, Memoris XVII Conferenica Internacional sobre Palma de Aceite, pp. 185–192 Tomo X.

Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R.L., Mouazen, A.M., 2016. Wheat yield prediction using machine learning and advanced sensing techniques. Comput. Electron. Agric. 121, 57–65. http://dx.doi.org/10.1016/j.compag.2015.11.018.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers Inc., San Francisco.

Pushparani, M., Sagaya, A., Ravan, S., 2018. Big data analytics using weight estimation algorithm for oil palm plantation domain. Int. J. Adv. Soft Comput. Appl. 10, 71–89.

Riedmiller, M., Braun, H., 1993. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: IEEE International Conference on Neural Networks, San Francisco, CA, pp. 586–591. 10.1109/ICNN.1993.298623.

Rowley, J., 2007. The wisdom hierarchy: representations of the DIKW hierarchy. J. Inf. Sci. 33, 163–180. http://dx.doi.org/10.1177/0165551506070706.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Networks 61, 85–117. http://dx.doi.org/10.1016/j.neunet.2014.09.003.

Scutari, M., 2010. Learning Bayesian networks with the bnlearn R package. J. Stat. Softw. 35, 1–22.

Scutari, M., 2017. Bayesian network constraint-based structure learning algorithms: parallel and optimized implementations in the bnlearn R package. J. Stat. Softw. 77, 1–20.

Tenenbaum, J.P., 1999. Bayesian modeling of human concept learning. In: Kearns, M.S., Solla, S.A., Cohn, D.A. (Eds.), Advances in Neural Information Processing Systems 11. MIT Press, Cambridge, MA.

USDA, 2018. United States. Foreign Agricultural Service. & United States. World Agricultural Outlook Board. Oilseeds, world markets and trade [electronic resource]/ United States Department of Agriculture, Foreign Agricultural Service.