Predicting *Prostephanus truncatus* (Horn) (Coleoptera: Bostrichidae) populations and associated grain damage in smallholder farmers' maize stores: A machine learning approach

Tinashe Nyabako[a], Brighton M. Mvumi[a]*, Tanya Stathers[b], Shaw Mlambo[a], Macdonald Mubayiwa[a]

[a]Soil Science & Agricultural Engineering, Faculty of Agriculture, University of Zimbabwe, Box MP 167 Harare, Zimbabwe; [b]Natural Resources Institute (NRI), University of Greenwich, Chatham Maritime, Kent ME4 4TB, UK

Abstract

Prostephanus truncatus is a notorious pest of stored-maize grain and its spread throughout sub-Saharan Africa has led to increased levels of grain storage losses. The current study developed models to predict the level of *P. truncatus* infestation and associated damage of maize grain in smallholder farmer stores. Data were gathered from grain storage trials conducted in Hwedza and Mbire districts of Zimbabwe and collated with weather data for each of the sites. Insect counts of *P. truncatus* and other common stored grain insect pests had a strong correlation with time of year with highest recorded numbers from January to May. Correlation analysis showed insect-generated grain dust from boring and feeding activity to be the best indicator of *P. truncatus* presence in stores (r = 0.70), while a moderate correlation (r = 0.48) was found between *P. truncatus* numbers and storage insect parasitic wasps, and grain damage levels significantly correlated with the presence of *Tribolium castaneum* (r=0.60). Models were developed for predicting *P. truncatus* infestation and grain damage using parameter selection algorithms and decision-tree machine learning algorithms with 10-fold cross-validation. The *P. truncatus* population size prediction model performance was weak (r = 0.43) due to the complicated sampling and detection of the pest and eight-week long period between sampling events. The grain damage prediction model had a stronger correlation coefficient (r = 0.93) and is a good estimator for *in situ* stored grain insect damage. The models were developed for use under southern Africa climatic conditions and can be improved with more input data for greater precision models to build decision-support tools for maize-based production systems.

Key words: Prediction model; insect grain damage prediction; decision tree; decision-support tools

1 Introduction

The prevailing climate has shifted leading to warmer temperatures, an increased frequency of drought and an increased occurrence of extreme events which pose a significant risk to the existing food and biological systems (Thornton *et al*., 2014). For example, the southern Africa region has experienced fluctuating rainfall patterns and increasing temperatures in the last two decades, with the semi-arid regions being the most vulnerable to extreme weather events and long dry periods (IPCC, 2014). Frequent droughts including impacts of *El Niño* reduce crop yields and increase the southern Africa region's food and nutrition insecurity risk status (FAO, 2018). The potential effects of global warming on pests can be explored based on knowledge of their physiological responses to specific weather factors using pest estimation models (Régnière, 2009; Maiorano *et al*., 2014). To-date, the effect of climate change and variability on grain storage management has been largely overlooked (Stathers *et al*., 2013; Moses *et al*., 2015), yet new ICTs and information gathering processes provide the possibility to create predictive early warning systems for storage pest management though challenges remain in gathering and packaging information from the field (Rashid, 2003; Wang *et al*., 2014).

Grain postharvest losses continue to threaten food security in sub-Saharan Africa (SSA) (Rembold *et al*., 2011). Emerging postharvest research and development studies regard the larger grain borer, *Prostephanus truncatus* (Horn) (Coleoptera: Bostrichidae) as a major threat to maize grain storage and thus food security across much of SSA (Mvumi & Stathers, 2014; APHLIS, 2018; Muatinte *et al*., 2019). The insect was accidentally introduced into Tanzania from its native central America at the end of the 1970's (Dunstan & Magazini, 1980; Hodges *et al*., 1983) and has now spread to most other African countries (Muatinte *et al*., 2014, 2019). The pest is known to infest maturing maize while it is still in the field (Giles & Leon, 1974) and persists throughout the subsequent postharvest stages including during the storage of shelled maize grain. Many pesticides are ineffective in controlling *P. truncatus* infestations in either the field or stored grain (Golob & Hanks, 1990; Mlambo *et al*., 2017, 2018). *Prostephanus truncatus* is also known to cause more than three times the damage of the normal spectrum of maize storage insect pests dominated by *Sitophilus zeamais* Motschulsky (Coleoptera: Curculionidae) (Makundi *et al*., 2010). The pest is largely spread through grain trade, in addition to its flight. Studies have shown that climate has an effect on the pest's food seeking flight behaviour (Borgemeister *et al*., 1998; Nansen *et al*., 2001) and that the pest also survives and breeds in forest habitats and wood (Nang'ayo *et al*., 1993; Muatinte & Van den Berg, 2019). The pest is known to have sporadic distribution patterns in natural environments and in stores (Krall, 1984; Birkinshaw *et al*., 2002; Boxall, 2002). A recent study using a correlative modelling tool, MaxEnt, explored which locations across the world would be climatically and ecologically suitable for the development of *P. truncatus* populations, and SSA was identified as a suitable host area (Arthur *et al.*, 2019). Most studies of *P. truncatus* have been concentrated in Eastern and Western Africa involving the sampling of commodities, evaluating the extent of infestations, and determining population ecology (Arthur *et al*, 2019). Though studies have been conducted to understand the behaviour of the pest in the natural ecosystem, models for predicting stored-grain infestation by the pest and the magnitude of its damage in smallholder stores are scarce in SSA.

A range of approaches exist for evaluating insect dynamics including regression, theoretical, non-parametric, phenology, and life-system models (Sharov, 1995). New approaches have been proposed to incorporate the modelling of ecological systems for improved agricultural management (Donatelli *et al*., 2017) with data-mining and development of pattern recognition as a plausible alternative (McQueen *et al*., 1995; Bhagawati *et al*., 2016), which can also be applied to postharvest-related data. Data-mining is derived from the ideas of statisticians, economists, forecasters, and communication engineers that patterns in data can be sought automatically, identified, validated, and used for prediction including in complex agricultural data (McQueen *et al*., 1995; Witten *et al*., 2016; Majumdar *et al*., 2017) such as storage insect dynamics. This can result in a better understanding of causes and effects of challenges such as crop production and postharvest pest occurrence, and can help inform agricultural decision-making (Gonzalez-Sanchez *et al*., 2014; Pham & Stack, 2018) in the face of changing agro-climatic conditions. Applying new scientific techniques and approaches to postharvest-related agricultural data can add value to the body of knowledge that currently exists and effectively allow better models to be developed (Moses *et al*., 2015) as IPM decision-support tools.

The term “machine learning” was first coined by pioneering computer gaming and artificial intelligence scientist Arthur Samuel in 1959. Machine learning refers to the automated detection of meaningful patterns in given data (Shalev-Shwartz & Ben-David, 2013; Sadiku *et al*., 2018). Machine learning is a more heuristic approach able to predict possible outcomes without the solution being necessarily optimal or perfect but offering a reliable solution to a problem when classic methods fail to come up with an exact solution (Witten *et al*., 2016). Supervised learning is used where we have prior knowledge of the output and is usually defined as a classification problem with the data containing categories, labels or classifications (Shalev-Shwartz & Ben-David, 2013; Witten *et al*., 2016) and is used to solve problems such as sorting and decision-making. In unsupervised learning on the other hand, the input data is not labelled or categorised so the learning process tries to find common traits in the data by which to cluster the data into subsets (Shalev-Shwartz & Ben-David, 2013; Witten *et al*., 2016). Unsupervised learning deals with clustering and association problems. Models derived from applying machine learning techniques can ultimately produce innovative software applications which are simple to use and can improve farm-level decision-making (Cunningham & Holmes, 1999; Patel & Patel, 2016; O’Grady & O’Hare, 2017) as decision support tools (Karim *et al*., 2017). Work on modelling insect pest damage in grain storage systems can build on systems approaches developed for field pest management (Teng & Savary, 1992; Donatelli *et al*., 2017).

The study objective was to develop a model for predicting *P. truncatus* numbers and insect grain damage in smallholder farmers’ grain stores using data collected from farmer-managed storage trials which were set-up by multi-stakeholder learning alliances (Mvumi *et al*., 2008), focused on building community resilience to climate-related risks through developing improved postharvest decision-support tools for better postharvest management.

2. Materials and methods

2.1 Description of study approach and grain storage sites

This study was part of two other experiments; one focussing on efficacy of synthetic storage pesticides (Mlambo *et al*., 2017; Mubayiwa *et al*., 2018) and the other on effectiveness of grain storage technologies (Mlambo *et al*., 2018); which were conducted concurrently over two grain storage seasons, *viz* August 2014 to May 2015 and August 2015 to May 2016. The experiments were conducted *in situ* and focused on collecting the grain insect pest profiles from non-pesticide treated maize grain stored in woven polypropylene bags to study the natural pest dynamics under different temperature and humidity regimes in the stored maize grain. Two districts, namely Hwedza and Mbire, were selected in Zimbabwe on the basis of their climate risk status in terms of flooding, temperature and rainfall change.

In Hwedza, two wards: Goneso and Makwarimba were selected. Goneso ward is located at a lower altitude (900-1200m) than Makwarimba (1200-1500m). Most of Makwarimba ward is in agro-ecological region IIb (Vincent *et al*., 1960) with annual rainfall of 750 – 1000 mm and mean annual temperature ranges of 18-30ºC. Goneso ward is in agro-ecological region III with mean annual temperatures of 18-35ºC and 650-800 mm annual rainfall (Vincent *et al*., 1960) (Figure 1). In Mbire, the two selected wards, namely Ward 8 and Ward 15, lie along the Zambezi valley at 500m above sea level in agro-ecological region IV (Vincent *et al*., 1960) with mean annual rainfall of 650-700 mm which usually falls within a 100-day period resulting in high flood incidence. Temperature averages 25ºC annually with summer temperatures reaching over 40ºC (Fritz *et al*., 2003).

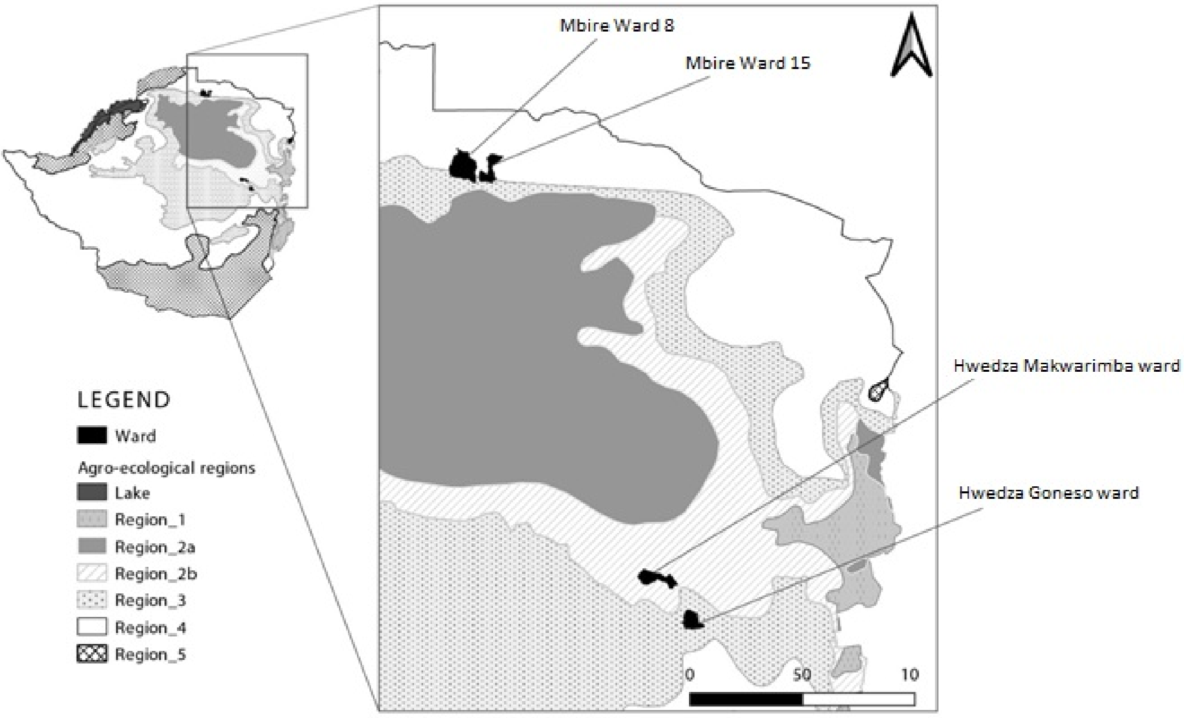

Figure 1: Location of sites used for the grain storage experiments in Mbire and Hwedza districts of Zimbabwe (Regions represent agro-ecological zones as described by Vincent *et al.* (1960))

The study was linked to on-going storage trials (Mlambo et al., 2017, 2018; Mubayiwa *et al*., 2018) which used a multi-stakeholder learning centre approach as described by Mashavave *et al*., (2011). Community leaders and local government extension workers assisted the research team in selecting the host-farmers. These stakeholders also worked together in setting-up the experiments and in monitoring and evaluation of the storage treatments. While the research team conducted eight-weekly sampling, laboratory grain damage, weight loss and insect species analysis and recorded the experimental data, the host farmers were responsible for maintaining the experimental environment between sampling events and noting any changes observed during the course of the experiments, and sharing information about the trials with neighbouring farmers. The learning centre approach aims to build local ownership of applied research to aid the integration of knowledge and technologies generated through the research into the local agricultural innovation systems. The host-farmers' sites were selected based on their accessibility, storage structure integrity, and security against theft.

2.2 Experiments, grain sampling and sample analysis

Grain procurement, trial-setting methodologies and stores are detailed in Mlambo *et al.* (2017, 2018) and Mubayiwa *et al*. (2018). Sampling of the stored grain was conducted at eight-week intervals over an eight-month period in the storage technology trials and a 10-month period in the synthetic pesticides efficacy experiments, coinciding with the length of time most farmers store their grain (Mvumi *et al*., 2003). The samples were analysed for insect numbers per species, damage, weight loss and dust from insect activity as described by Mlambo *et al.* (2017, 2018) and Mubayiwa *et al*. (2018). Extech Instuments® Humidity / Temperature Dataloggers Model RHT10 (FLIR Systems, Inc., Nashua, U.S.A) were installed under the roofs of selected representative storage facilities to measure store temperature and humidity at 30-minute intervals from September 2014 to April 2015 and from August 2015 to May 2016. The data were downloaded and saved at bi-monthly intervals.

The sampling process included feedback on the state of the grain using the various protectants and the use of wire and bead models for easy pest identification by the farmers. The samples were sieved to separate the dust and insects from the grain as described by Mlambo *et al.* (2017, 2018) and Mubayiwa *et al*. (2018). For the purpose of this study, a total of 13 experimental variables were recorded from the untreated maize grain samples and coded as follows:

1. Mc (moisture) - Grain moisture content (%)
2. Tmean (temperature) - Store temperature (°C)
3. RHmean (relative_humidity) - Store relative humidity (%rh)
4. Wk (week)– Numbered week of the year: Storage season generally begins May 1st in Zimbabwe (week 18) of the year starting January
5. C (dust)- Dust content from insect feeding/kg sample (%)
6. D (damaged grains) - Damaged grains/kg sample (%)
7. R (rotten grains) - Rotten grains/kg sample(%)
8. Sz (sitophilus) - Number of adult *Sitophilus zeamais* insects/kg sample
9. Tc (tibolium) - Number of adult *Tribolium castaneum* insects/kg sample
10. Sc (sitotroga) - Number of adult *Sitotroga cerealella* insects/kg sample
11. *P. truncatus* (lgb) - Number of adult *Prostephanus truncatus* insects/kg sample
12. Rd (rhyzopertha) - Number of adult *Rhyzopertha dominica* insects/kg sample
13. Wa (wasps) - Number of adult parasitic wasps of the order hymenoptera/kg sample

Dust refers to all grain dust produced as a result of insect feeding and boring, and insect exuviae. The number of insects included both the dead and live adult insects since the damage and dust recorded at sampling is a product of the feeding habits of insects prior to death, live insects and residual dead insect matter between sampling dates (Makundi *et al*., 2010). The wasps were those which parasitise grain storage insects of the Anthoricidae family and Pteromalidae family such as *Pteromalus cerealellae* and *Anisopteromalus calandrae* which can attack larvae of primary pests including *P. truncatus* (Hodges *et al*., 1983; Savidan, 2002; Bonu-Ire *et al*., 2015; CAB International, 2018). However, for purposes of this study, the different parasitic wasp species were not distinguished.

2.3 Hyperparameter optimisation and parameter selection for *P. truncatus* and grain damage

The Waikato Environment for Knowledge Analysis (WEKA® version 3.8.2) software (Hall *et al.*, 2009; Frank *et al.*, 2016) was selected for use for data analysis because it is a proven data-mining and machine learning platform and is a free and open source software (FOSS) based on the equally free Java programming language (David *et al.*, 2013; Sharma *et al.*, 2015; Kotthoff *et al.*, 2017; Witten *et al.*, 2017).

The AutoWEKA algorithm was used to perform hyperparameter optimisation using Bayesian Optimisation to find a strong instantiation of a dataset (Thornton *et al*., 2013). It considers the combined space of WEKA's learning algorithms A = {$A^{(1)}, \ldots, A^{(k)}$} and their associated hyperparameter spaces $A^{(1)}, \ldots, A^{(k)}$ and aims to identify the combination of algorithm $A^{(j)} \in$ A and hyperparameters $\lambda \in \Lambda^{(j)}$ that minimises cross-validation loss

$$A^*_{\lambda*} \epsilon \; argmin^{argmin}_{A(j)\in A,\lambda\in\Lambda(j)} = \frac{1}{k}\sum_{i=1}^{k} L(A^{(i)}_{\lambda}, D^{(i)}_{train}, D^{(i)}_{test})$$

Where $L(A^{(i)}_{\lambda}, D^{(i)}_{train}, D^{(i)}_{test})$ denotes the loss achieved by algorithm A with hyperparameters λ when trained on D_{train} and evaluated on D_{test} (Thornton *et al*., 2013; Kotthoff *et al*., 2017). As the number of possible algorithms that could have been used in developing the models are vast, Bayesian optimisation procedures, Sequential model-based algorithm configuration (SMAC) and Tree-structure Parzen Estimator (TPE) were used to find combinations of algorithms and hyperparameters that often outperform existing baseline methods (Thornton *et al*., 2013; Kotthoff *et al*., 2017). Algorithms considered included decision-trees, k-nearest neighbours, multi-layer perception, support vector machines and linear regression.

The CfsSubsetEval algorithm was suggested in AutoWEKA for parameter selection as it evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with a degree of redundancy between them; thus only using features that maximise accuracy (McQueen *et al*., 1995; Hall, 1999; Kuhn & Johnson, 2013). The search method for the CfsSubsetEval algorithm parameter selection process was the Best-First which is a method that does not just terminate when the performance starts to drop but keeps a list of all attribute subsets evaluated so far, sorted in order of the performance measure, so that it can revisit an earlier configuration instead (Witten *et al*., 2016).

Regression-based algorithms were purposefully considered to develop the preferred output regression-based models for academic purposes as they work with numeric prediction as opposed to non-regression-based models produced by algorithms such as Random Forests (Witten *et al*., 2016). Linear Regression and decision-tree induction were ultimately selected from the suggested models from running the AutoWEKA algorithm on the dataset as the best-fit for developing the models as they have the following properties:

1. Linear regression – $Y_{i=} \beta_0 + \beta_i X_i$
   Where Yi is the dependent variable, β_0 is the intercept, β_i is the slope and X_i are the n observations of the independent variable (Rawlings *et al*., 1998). This method expresses the class as a linear combination of the attributes (Witten *et al*., 2016) and was used to develop a linear equation model for both *P. truncatus* and grain damage from the 13 afore-mentioned variables.
2. Decision-tree induction – a decision-tree is a flowchart-like tree structure, where each internal node denotes a test on an attribute, each branch represents a result of the test, and each leaf node holds a class label as shown in Figure 2 (Wang & Witten, 1996; Frank *et al*., 1998; Grabczewski, 2014; Barros *et al*., 2015; Sharma *et al*., 2015). The pruned M5 model tree learning algorithm (implemented as M5P in WEKA®) was applied on the data to predict damage and *P. truncatus* count as it is ideal for use with numeric data for numeric predictions (Onyari & Ilunga, 2013; Sharma *et al*., 2015).

Pruning reduces the size of the decision-tree by removing redundant sections, which helps in reducing the complexity of the final tree, thereby improving predictive accuracy (Mansour, 1997; Witten *et al*., 2016).

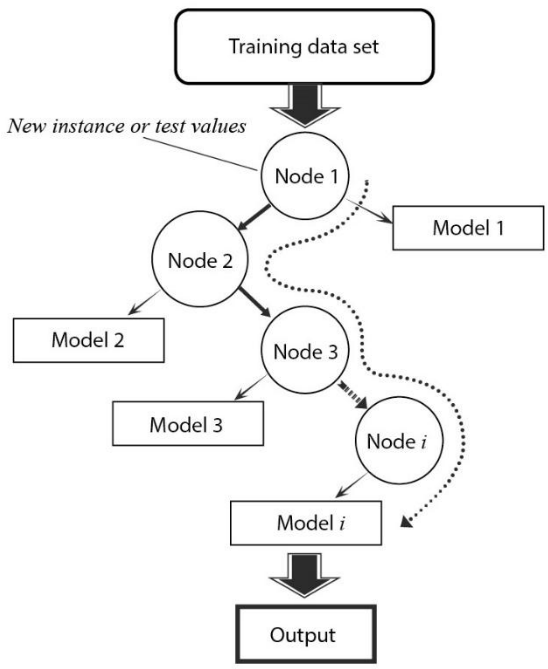

Figure 2: A pruned decision-tree (Adapted from Witten *et al.* (2016))

2.4 Data exploration and development of prediction models

2.4.1 General approach

The data containing selected parameters were iteratively resolved into a three cluster classification (Abernethy, 2010) using WEKA®. Clustering was performed using the K-Means algorithm with the advantage being that it divides a dataset into clusters (groups of data points that belong together) where each cluster has points which are similar to each other (Abernethy, 2010; Trevino, 2019). Visualisation of the clustered data using Python® software with Pandas®, Matplotlib® and Seaborn® among the Python modules used (Microsoft, 2016; edX, 2018; Mendis, 2019). A Jupyter Notebook® was used to perform the visualisations including kernel density estimation (Winner, 1985) and distribution plots used to plot distribution of results as a way of observing patterns in the data. Due to the numeric nature of the required prediction, regression modelling options were considered. The machine learning approach was eventually selected due to the heuristic approach it offers which allows data-mining, parameter selection and hyperparameter optimisation (Chapelle *et al.*, 2001; Wahbeh *et al.*, 2011; Thornton *et al.*, 2013; Witten *et al.*, 2017). Machine learning is also better suited for finding patterns on complex data with many variables or smaller datasets (Witten *et al.*, 2016). In this experiment, 13 variables were recorded at bi-monthly intervals; hence the dataset was small and the number of variables fairly high.

2.4.2 Developing a model for *P. truncatus* count and grain damage

Stochastic and deterministic approaches were considered for developing the model (Soetaert & Herman, 2009; Tonnang *et al*., 2017) with the stochastic approach being ultimately chosen as it requires fewer assumptions and has limited overfitting (Witten *et al*., 2016). Machine learning algorithms can also be modified to clean the partially clean data and deal with missing values (Witten *et al*., 2016). An unsupervised learning approach (Wahbeh *et al*., 2011; Witten *et al*., 2016) was used to infer the natural structure in the dataset as related to *P. truncatus* count and grain damage as the selected target parameters. In general, an instance is a single record in a dataset characterized by the values of features, or attributes, that measure different aspects of the instance (Witten *et al*., 2016) which in this case, consisted of the ±13 records per instance. A total of 186 instances used to develop the model were collected over two storage seasons (2014/15 and 2015/2016).

2.4.3 Model validation

The preferred validation method was k-fold cross-validation with k = 10 as it reduces over-fitting as compared to random sub-sampling and the holdout method especially for smaller datasets (Blockeel & Struyf, 2001; Witten *et al*., 2016) as in our case with 189 instances of data collected (Figure 3). This validation method is heavy and requires adequate computing power (Witten *et al*., 2016). The 10-fold selection is based on theoretical evidence and extensive tests on numerous datasets with different machine learning techniques which gives the best estimate of error (Witten *et al*., 2016). The data were divided randomly into 10 parts in which the class is represented in approximately the same proportions as in the full dataset with each part held out in turn and the learning scheme trained on the remaining nine-tenths; then its error rate calculated on the holdout set (Kuhn & Johnson, 2013; Witten *et al*., 2016).

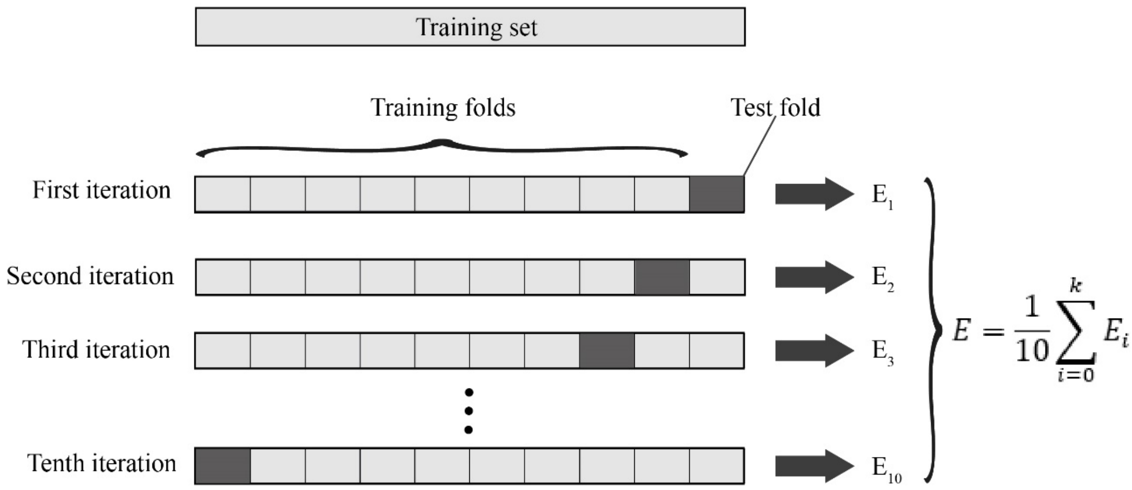

Figure 3: K-fold validation process with k = 10

An iterative process was used with Linear regression and M5P algorithms under 10-fold validation to produce solutions from which the model that maximised the correlation coefficient and minimised errors (Mean-squared error, Root mean-squared error, Mean absolute error, Relative squared error, Root relative squared error and Relative absolute error) was selected (Witten *et al*., 2016). The resulting models were stored in Java-based model files which are useable with any Java programming language Integrated Development Environments (IDEs) including Android Studio®.

2.4.4 Clustering for similar instances

A K-Means algorithm was used to create a classes-to-clusters evaluation to find the minimum error mapping of the classes in our data to clusters (only the class labels that correspond to the instances in a cluster are considered for each cluster), with the constraint that a class can only be mapped to one cluster (FutureLearn, 2019).

Clustering was implemented in WEKA® and used to visualise the collected data to show the separation of classes in the data and give an indicator of the sources of error in the classification algorithms (Faith, 2007) (Figure 4).

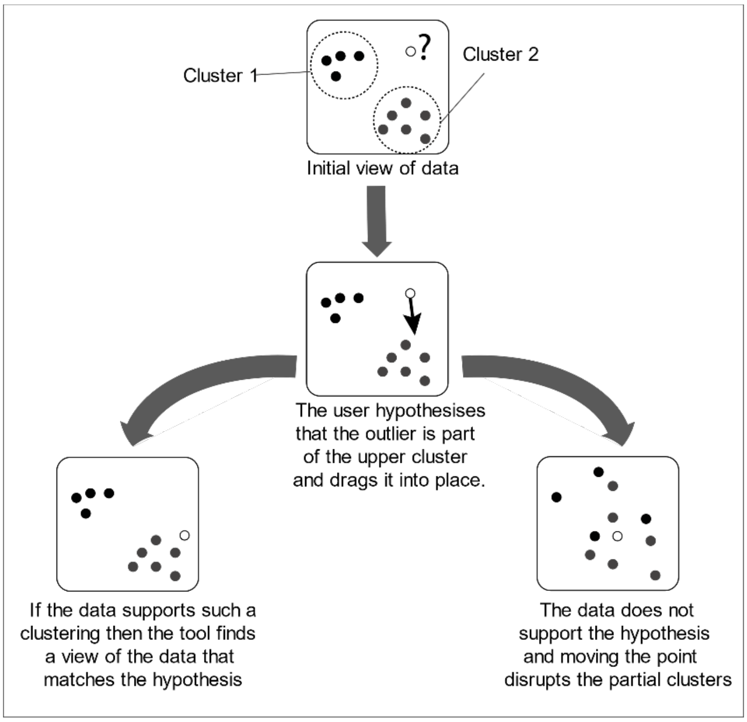

Figure 4: The use of targeted projection pursuit for interactive data exploration (Adapted from Faith, 2007)

Techniques based on linear projections have the advantage of not only showing an informative view of the data, but the weights of the projection itself which may include useful information (Faith, 2007). For example, if one particular projection is found to show a clear separation between classes in the data, then the most significant weights in the underlying projection will indicate which variables in the original data were the best discriminators for those classes (Faith, 2007).

3. Results

3.1 Exploring the selected data for *P. truncatus* prediction

After conducting a parameter selection process on the source data the most important parameters influencing *P. truncatus* count and grain damage were determined as summarised in Table 1.

Table 1: WEKA® -selected attributes for predicting *P. truncatus* numbers and damage using CfsSubsetEval and BestFirst search

Selected parameters for predicting *P. truncatus* numbers	Selected parameters for predicting % grain damage
dust (%)	moisture content (%)
Wasps (number)	temperature (ºC)
Tribolium (number)	storage time (weeks)
Relative humidity	dust (%)
	rotten (%)
	Tribolium (number)
	Sitotroga (number)
	Rhyzopertha (number)
	Wasps (number)

The selected data were clustered using K-Means into three groups as shown by the confusion matrix in Table 2.

Table 2: Final cluster centroids and confusion matrix of selected parameters

		Cluster No.		
Attribute **No. of instances**	**Full Data** **(159)**	**0** **(43)**	**1** **(61)**	**2** **(55)**
Moisture	10.71	11.45	10.40	10.49
Temperature	26.59	23.84	26.59	28.74
Relative humidity	53.05	56.21	53.05	50.59
Week	24.94	19.21	42.20	10.27
Dust	1.92	6.11	0.12	0.65
Damage	30.75	72.40	4.58	27.20
Rotten	1.09	0.63	1.34	1.18
Sitophilus	32.77	57.47	11.10	37.48
Tribolium	17.84	45.76	0.64	15.08
Sitotroga	55.77	129.50	22.91	34.56
Lgb(*Prostephanus)*	22.70	79.31	0.64	2.89
Rhyzopertha	0.73	2.16	0.04	0.38
Wasps	1.45	4.30	0.38	0.40

Confusion matrix

		0	***1***	***2***
		41	0	0
		0	61	0
	0	1	0	55
Clustered instances		***25.8%***	***38.4%***	***34.6%***
Incorrectly clustered instances : 2.0	*1.3 %*			

Temperatures were observed to be high (between 30°C and 37°C) during the second and third quarters of the year from August to December. The density plots of the clusters according to the main parameters of interest are shown in Figure 5. Analysis of weather parameters and *P. truncatus* counts showed the highest counts at relative humidity between 55% and 70%. Grain damage levels peaked between February and May (Weeks 5 – 18).

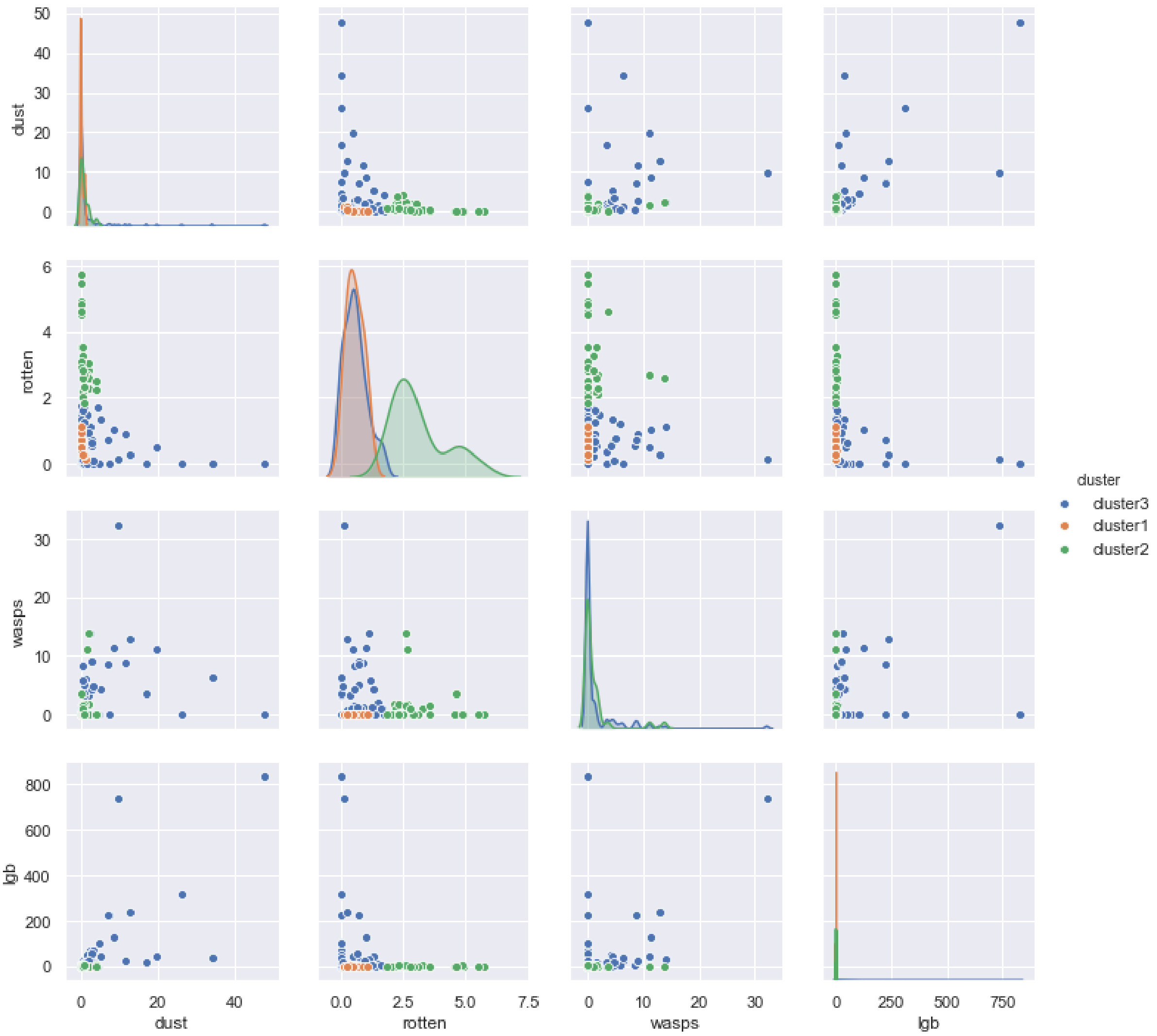

Figure 5: Scatter and kernel density plots of the selected data with clusters (lgb = *Prostephanus truncatus*)

A violin plot was used to show the distribution of the data in the clusters revealing cluster 3 and cluster 1 closely largely overlaying while cluster 2 appears largely distinct from the other two (Figure 6).

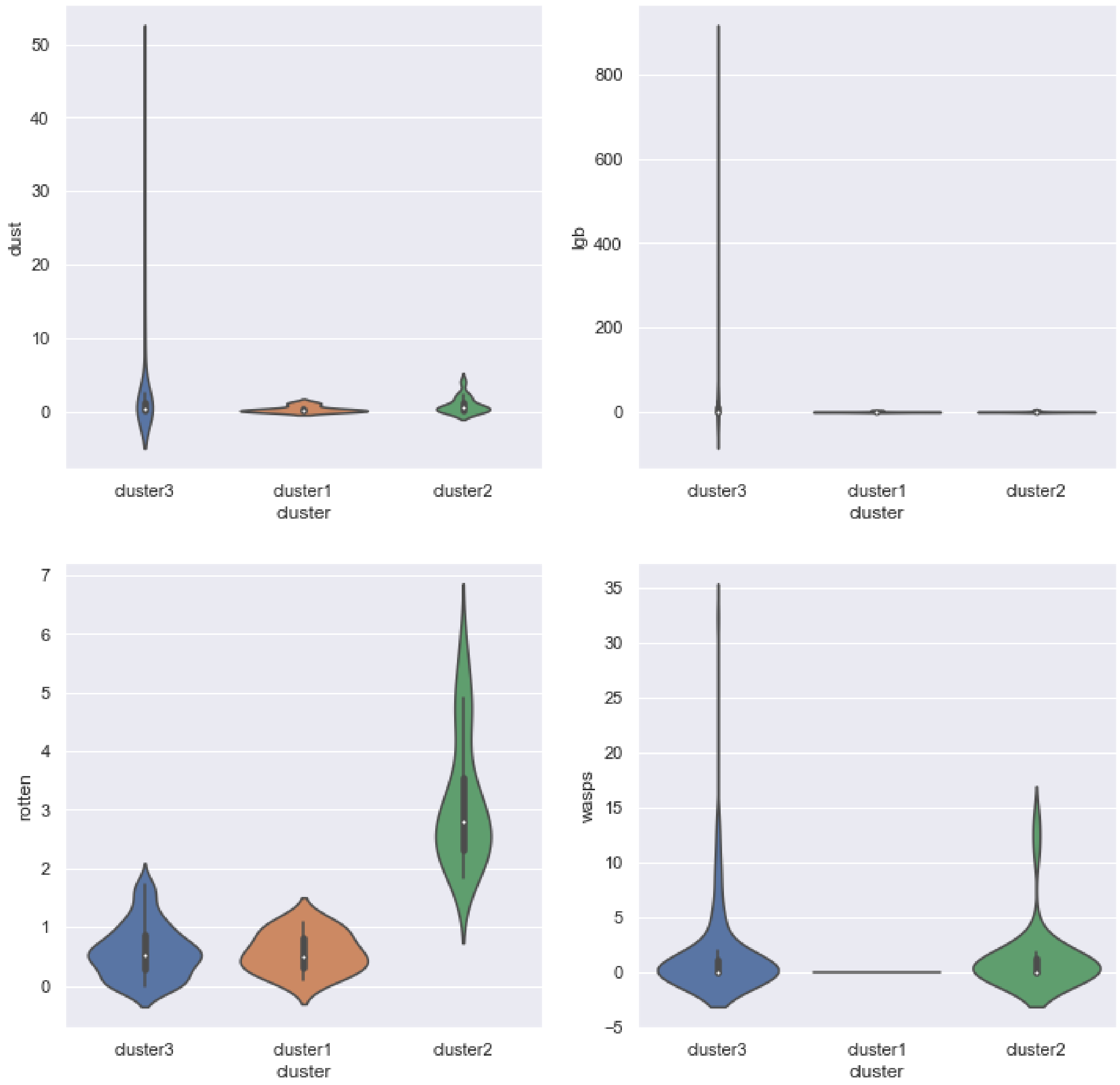

Figure 6: Violin plot showing distribution of data points for different variables within clusters

A correlation matrix was plotted to highlight the correlation between the different variables and *P. truncatus* distribution (Table 3). Notable correlation was between *P. truncatus* (lgb) and dust (r=0.70), and between lgb and wasps (r=0.48).

Table 3: Pearson correlation values of the collected variables for predicting *P. truncatus* numbers and insect grain damage

	Moisture	**Temperature**	**Humidity**	**Week**	**Dust**	**Damaged**	**Rotten**	**Sitophilus**	**Tribolium**	**Sitotroga**	**Lgb**	**Rhyzopertha**	**Parasitic wasps**
Moisture	1.00*	-0.63*	0.60*	-0.20*	0.27*	0.37*	-0.20*	0.28*	0.16*	0.11	0.15	0.02	0.13
Temperature		1.00*	-0.74*	-0.11	-0.34*	-0.44*	-0.09	-0.18	-0.2	-0.28*	-0.32*	-0.02	-0.19
Humidity			1.00*	-0.04	0.26*	0.38*	0.13	0.16	0.26*	0.09	0.29*	0.02	0.13
Week				1.00*	-0.14	-0.49*	0.1	-0.24*	-0.28*	-0.15	-0.08	-0.14	-0.15
Dust					1.00*	0.51*	-0.17*	0.34*	0.58*	-0.06	0.70*	0.05	0.30*
Damaged						1.00*	-0.35*	0.28*	0.60*	0.44*	0.41*	0.21*	0.36*
Rotten							1.00*	0.13	-0.28*	-0.24*	-0.17*	-0.08	-0.05
Sitophilus								1.00*	0.05	-0.06	0.06	-0.02	0.28*
Tribolium									1.00*	0.11	0.54*	0.19*	0.23*
Sitotroga										1.00*	0.06	-0.07	0.09
Lgb											1.00*	-0.01	0.48*
Rhyzopertha												1.00*	0.16*
Parasitic Wasps													1.00*

*Note: * represents a correlation value with a p-value < 0.05.*

3.2 *Prostephanus truncatus* model after parameter selection

The output of evaluation of the decision-tree model for predicting *P. truncatus* is shown in Table 4. The correlation coefficient is low at 0.43 with large root mean square error of 81.93.

Table 4: Model fit results for predicting number of *P. truncatus*

M5P decision tree model statistics	
Correlation coefficient	0.43
Kendall's tau	0.40
Mean absolute percentage error	∞
Root mean square percentage error	∞
Spearman's rho	0.51
Mean absolute error	23.93
Root mean squared error	81.93
Relative absolute error	76.83 %
Root relative squared error	92.32 %
Total Number of Instances	189

The decision-tree regression model for *P. truncatus* was calculated using the M5P algorithm as shown in equation 1:

P. truncatus model;

No. of *P. truncatus* = 10.4663 * C + 7.3264 * Wa - 8.2842

Where: C (dust) = Dust content (%) /kg sample

Wa = Number of adult parasitic wasps of the order hymenoptera /kg sample

3.3 Insect grain damage model after parameter selection

Parameter selection for insect grain damage yielded nine parameters which were then used as grain damage predictors. The recorded data were further clustered using the K-Means algorithm. A kernel density estimation plot and clustered scatter plot was generated (Figure 7).

Figure 7: Clustered kernel density and scatter plots showing the relationship between grain damage and climate related variables

High levels of grain damage above 20% were observed in samples starting from weeks 5 to 35 following moisture content rise to above 12% and average temperatures of between 25°C and 35°C over the same period. Grain damage peaked when grain moisture content was between 11% and 15% (Figure 7).

The general trend showed that increased grain moisture content coincided with higher *P. truncatus* numbers and higher damage between March and May (weeks 9 to 18). Pest numbers were generally higher during the early weeks of the year when rainfall and warm to hot temperatures dominated (Figure 6). Insect numbers were notably high for *T. castaneum* and *S. cerealella* during this period.

In the violin plots for parameters selected as predictors of grain damage. week, temperature and moisture showed visually different centroid points (Fig 8).

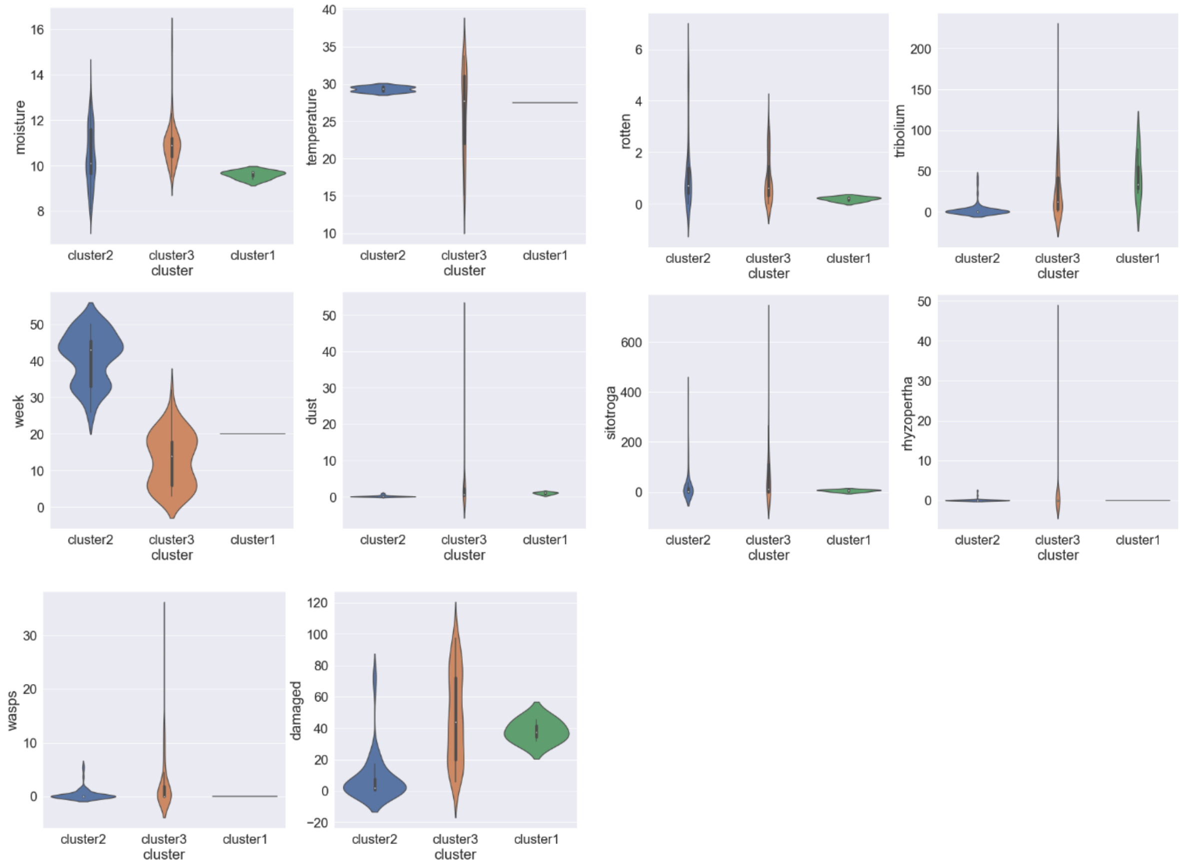

Figure 8: Violin plot showing distribution of clustered data points for the 10 parameters selected

Cluster 2 showed the greatest deviation in centroids from clusters 1 and 3 for parameters week, damaged, and dust (Figure 8). Clusters 1 and 2 had the least deviation from its centroid for dust and insect species. The M5P algorithm in WEKA® was used with tree pruning to remove tree branches with less than 10 instances and 10-fold cross validation.

3.4 Modelling grain damage

Using the decision tree M5P algorithm in WEKA®, a high correlation coefficient of 0.9288 was achieved with a low root mean square error of 10.77 (Table 5).

Table 5: Model fit 10-fold cross validation summary for insect grain damage

M5P decision tree model statistics	
Correlation coefficient	0.93
Kendall's tau	0.75
Mean absolute percentage error	∞
Root mean square percentage error	∞
Spearman's rho	0.91
Mean absolute error	6.99
Root mean squared error	10.77
Relative absolute error	28.29 %
Root relative squared error	37.09 %
Total Number of Instances	186
Ignored Class Unknown Instances	3

The decision-tree illustration produced by the M5P algorithm for predicting grain damage from which four equations were derived is shown in Figure 9.

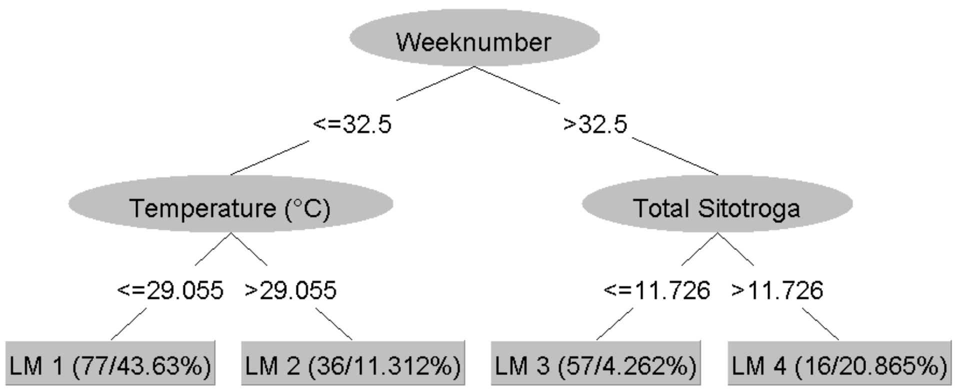

Figure 9: Decision-tree classifier model for damage prediction after parameter selection

Note: LM means Linear Model. The format (77/43.63%) indicates that 43.63 % of the data set instances were correctly classified/predicted along the named branch of the decision-tree. The <=32.5 denotes "if weeknumber <=32.5 then temperature becomes the next most important decision parameter" etc.

M5P tree model produced the following equations for % insect grain damage are shown in Equation 2.

<u>*Grain damage model;*</u>

Number of rules = 4

LM num: 1

D = 3.7781 * Mc - 0.328 * Tmean + 0.5342 * Wk + 1.0326 * C - 8.8884 * R + 0.0235 * Tc + 0.0489 * Ts + 0.6425 * Rd + 0.2108 * Wa + 11.6829

LM num: 2

D = 1.8645 * Mc + 0.1239 * Tmean + 0.2576 * Wk + 10.8734 * C - 2.5159 * R + 0.0235 * Tc + 0.1224 * Ts + 0.3174 * Rd + 0.2859 * Wa - 11.3511

LM num: 3

D = 0.0991 * Mc - 0.1904 * Tmean - 0.1101 * Wk + 12.2993 * C - 0.5149 * R + 0.0342 * Tc + 0.1303 * Ts + 0.1433 * Rd + 0.1707 * Wa + 11.3155

LM num: 4

D = 0.3968 * Mc - 0.1904 * Tmean - 0.1417 * Wk + 7.5753 * C + 3.8116 * R + 0.0342 * Tc + 0.0376 * Ts + 0.1433 * Rd + 0.1707 * Wa + 11.2657

The model parameters are as described in Section 2.2.

4. Discussion

The ambient temperatures recorded in the two focal wards in Mbire district, and in Goneso ward of Hwedza district were higher than those in Makwarimba ward of Hwedza district. Previous work suggests that temperature can influence pest populations (Worner, 1998; Munyuri & Tabu, 2013). *Prostephanus truncatus* is present in both Makwarimba and Goneso wards of Hwedza district, with higher pest numbers and higher mean annual temperatures in Goneso than in Makwarimba ward (Mlambo *et al*., 2017). In previous modelling work, parameters such as storage duration, temperature, humidity, grain moisture content, and the developmental stages of *P. truncatus* have been shown to influence numbers of the pest under smallholder farm conditions (Meikle *et al*., 1998).

Adult *P. truncatus* do not favour high grain dust situations and tend to fly away in search of alternative food sources (Borgemeister *et al*., 1998; Nansen *et al*., 2001) leading to higher numbers of the other grain insect pests such as *T. castaneum* and *S. cerealella* being recorded. This explains our results in the second half of the storage season when insect grain damage and dust quantities had increased. In laboratory experiments conducted with *P. truncatus* and *S. zeamais*, it was observed that the latter was more competitive at lower temperature while the former was responsible for more damage and produced more progeny as temperature rose above 30°C to 35°C (Quellhorst et al., 2019). Similar results were observed in experiments involving *P. truncatus* and *R. dominica* with a further observation that that in an enclosure, *P. truncatus* outcompetes *R. dominica* in the 30°C to 35°C temperature range, more likely due to *P. truncatus* having comparatively higher preference for maize grain as a food source than direct competition (Sakka and Athanassiou, 2018). This has implications for maize grain storage in areas with generally high annual temperatures as the threat of *P. truncatus* becomes more pronounced.

As the annual rains fall in the summer months in Zimbabwe from October to April, *P. truncatus* numbers generally increase as temperatures approach the pest's optimal developmental temperature of 32ºC and 80% relative humidity (Shires, 1979). Hence, *P. truncatus* populations increased together with other insect pests as the rainfall season commenced and as summer temperatures increased; a trend also observed by Stathers *et al.* (2008). The high temperatures experienced during the dry third quarter of the year into the hot and humid fourth quarter and the first quarter of the following year, provide adequate conditions for *P. truncatus* larval survival with the result being a strong presence of the pest in the second quarter of the subsequent year as was also found by Meikle *et al.* (1998). Flight activity of *P. truncatus* increases in the temperature range from 20 to 30°C but declines sharply once it reaches 35°C (Fadamiro & Wyatt, 1995). This suggests that there may be less movement of the pest from stored grain during the peak dry temperatures observed in the second half of the year thus contributing to the high damage later observed as the postharvest season draws to an end in the following year. Grain damage is an important parameter as damage caused by *P. truncatus* can lead to lower consumer valuation and greater price discounts than damage by other storage insect pests such as *Sitophilus* spp. (Boxall, 2002). The relationship between moisture content and relative humidity with *P. truncatus* numbers also agrees with findings by (Meikle *et al*., 1998). In the current study, grain damage levels were observed to be highest at temperatures between 15°C and 30°C and relative humidities between 45% and 75%. The high damage observed towards the cooler winter was a result of cumulative damage during the storage period when summer temperatures and rains in the first half of the year provided favourable conditions for the storage pests to flourish. This also coincides with the *P. truncatus* populations, though total insect grain damage is actually a product of feeding by multiple species of stored product insects including *Sitophilus* spp., *S. cerealella* and *Tribolium* spp.

Projections of future maize grain output and variation of temperature and precipitation in Zimbabwe suggest an increase in areas having an average annual temperature above 25°C and an increase in areas receiving annual precipitation below 610 mm (Nyabako & Manzungu, 2012). These changes in weather patterns may aid more rapid build-up of *P. truncatus* populations including in the surrounding natural environment where it thrives in some indigenous tree and shrub species (Nang'ayo et al., 1993) as suggested in other studies (Arthur et al., 2019). It should be noted, however, that only the male-produced pheromone influences *P. truncatus* dispersal and host-finding behaviour, with food volatiles having no effect on the host selection (Fadamiro et al., 1998). High total *P. truncatus* numbers were recorded at temperatures of between 15°C and 30°C which may be a result of high numbers of dead *P. truncatus* insects being recorded from March to May (see details in Mlambo et al. 2017; 2018) which coincides with the start of low winter temperatures in the southern African region. It is possible that adult *P. truncatus* were attracted to the pheromones of insects in infested grain from neighbouring grain stores within the community or the natural environment. It would be informative to also investigate the *P. truncatus* population dynamics outside the store to determine the relationship between incoming versus resident (in-store) infestation and subsequent insect grain damage. Boring of grain storage bags by *P. truncatus* may be a result of the pest movement between the bags of stored grain and the natural environment. The pest causes extensive damage to the integrity of some bags as well as to the grain; resulting in generation of copious amounts of grain dust during boring and feeding. *P. truncatus* grain damage can be identified by the presence of circular holes on the polypropylene bags with a lot of grain dust trickling from these holes. *P. truncatus* grain damage often results in in structural integrity failure of the bags and ripping apart while being moved or sampled (personal experience) as also reported in earlier studies (Nang'ayo et al., 1993; Birkinshaw et al., 2002; Hodges, 2002).

Grain damage peaked together with both *P. truncatus* and wasp counts between February and June (weeks 5 to 22) as the postharvest season ended. The correlations between dust and *P. truncatus* (r = 0.70), and between dust and insect grain damage (r = 0.51) were significant, demonstrating the strength of using dust as a visual indicator of the presence of *P. truncatus* and grain damage which affect grain quality. Dust content and grain weight loss are a good indication of increase of insect feeding activity and grain damage, and that is why these parameters are measured when conducting loss assessments of grain.

Among the different insect species, the highest correlations were observed between *P. truncatus* and wasp numbers (r = 0.48), and between *P. truncatus* and *T. castaneum* (r=0.54). Observations indicate that development of *P. truncatus* is partially inhibited by *T. castaneum* and *S. zeamais* (Kenneth, 1988). *Sitophilus zeamais* larvae are known to deter *P. truncatus* from infesting maize as *P. truncatus* prefers uninfested grains than grains that have already been infested (Danho *et al*., 2000). *Prostephanus truncatus*, being a primary pest, favours undamaged grains and is known to produce a lot of dust, which sustains secondary pests such as *T. castaneum* which prefer to feed on damaged grains (Hodges, 1986). Hence *P. truncatus* feeding activity may lead to increases in occurrence of *T. castaneum* as was also found by (Mlambo *et al*., 2017).

The number of parasitic wasps and *P. truncatus* beetles in the untreated grain were moderately correlated (r = 0.48) which explains why the model for *P. truncatus* had wasps as a major parameter.

Wasps tend to be highly susceptible to chemical grain treatments (Perez-Mendoza *et al*., 1999) and hence are usually only present in untreated grain. The small size and high mobility of the wasps make them much more difficult to count when live than the insect pests, implying a more thorough sampling and counting mechanism for *P. truncatus* and wasps may be required in future experiments. Where temperatures are relatively high above 25℃ as is the case in Mbire, the survival of wasps may be negatively affected compared to the host pest which has been shown to favour generally higher temperatures of above 35℃ in laboratory experiments (Shires, 1979). Increased ambient moisture availability during the rainy season generally increases insect feeding and grain damage by both primary and secondary pests. Rotting does not seem to have much effect on *P. truncatus* numbers though it does have some effect on damage level which may be due to the feeding of other insect pest species within the grain.

Machine learning techniques were used to iteratively develop the models for predicting *P. truncatus* and grain damage as they could provide a reliable model considering the low number of data instances with *k*-fold cross-validation used for validation of the model (Blockeel & Struyf, 2001; Witten *et al*., 2016). Parameter selection was applied to determine the parameters which should be included in the models. The prediction models for *P. truncatus* and grain damage were processed iteratively using the decision-tree algorithm M5P in WEKA® software. The prediction model for *P. truncatus* had a correlation coefficient of 0.43 which is low but can be attributed to the complexity in sampling for *P. truncatus* and the nature of the experimental *in situ* environment that the experiments were conducted in. Further challenges to the model accuracy could have arisen from the low sampling frequency of eight weeks which may have affected analysis of *P. truncatus* activity and its correlation with the other variables. The accuracy of the model is, dependent on the quantity of data; with more data leading to better prediction accuracy (Witten *et al*., 2016); though it must also be acknowledged that frequent sampling for such *in situ* experiments comes with the risk of disturbing the ecosystem too often which can affect normal population development of the storage insects.

Insect grain damage prediction produced a high correlation coefficient ($r = 0.93$) which is an indication of the confidence we can have in using the model as an estimator of grain damage. The moisture content and wasp count were the greatest factors in predicting damage. It should be noted that we can use the relationship between parasitic wasps and *P. truncatus* count shown in the *P. truncatus* model to substitute for *P. truncatus* in the damage model. Ability to predict potential insect damage to grain in storage can aid decision-making in terms of the most appropriate grain protection method depending on intended storage period and intended use of the grain. The model can be used to build applications for estimating grain condition at different times of the year under different conditions and can ultimately contribute to development of more tools for farming stakeholders including agricultural extension agents to be able to provide quicker grain pest risk assessments in stored grain and advise on preparatory measures accordingly.

While more advanced systems of grain storage in industrialised countries offer a more stable, uniform and controlled storage environment for grain pest behavioural analysis, the smallholder farmers' maize stores tend to be more similar to a natural ecosystem, which results in greater immigration and emigration of *P. truncatus* in relation to the store; making it more difficult to model the behaviour of the pest (Meikle *et al*., 1998). Accurate sampling for *P. truncatus* in the smallholder grain stores was difficult due to the pest's feeding habits as the pest tends to aggregate at the base of grain bags or stacks or bulk storage deposits (Hodges, 2002; Hodges *et al*., 2003). Multi-compartmentalised grain probes were used for sampling the stored bagged maize grain, and they have apertures at intervals along the 1.5 m length of the probe where grain samples are obtained from different vertical positions within a grain bag or grain bulk. The probe was inserted repeatedly at different positions from the open top surface of the grain bag to provide the most spatially representative sample of grain. However, the design of the grain sampling probe is such that it has a conical tip to facilitate probing of the bagged grain, leaving a 5 cm space at the bottom which goes unsampled; hence *P. truncatus* present in the lowest part of a bag may not be sampled accurately (Chigoverah & Mvumi, 2018) as illustrated in Figure 10.

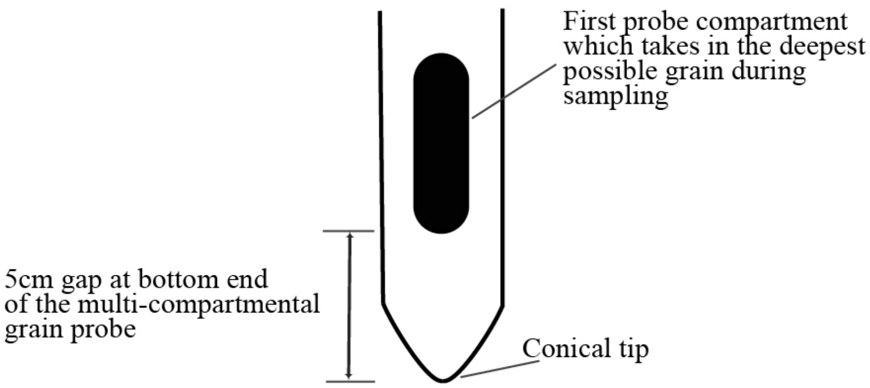

Figure 10: Bottom end of multi-compartmentalised grain sampling probe showing the bottom portion of the probe that fails to get the last 5cm of grain where *Prostephanus truncatus* is mostly found

This problem can be overcome by carefully emptying the grain from the different depth layers of a sack, store or bulk and taking samples from each of these depths, but this is hugely laborious and causes disturbance of developing insect populations and would affect subsequent samplings. Additionally, there is a further sampling problem with *P. truncatus* as the insect is an internal feeder of individual kernels (Holst *et al*., 2000); hence there can be detection errors during manual sample analysis as many adult *P. truncatus* may remain inside the grains despite sieving and may therefore not be recorded. The experiments were conducted in farmers' stores. Thus the feeding, flight and infestation activities of *P. truncatus* is potentially problematic to the farmers households due to its boring of timber, furniture, curtains and many other non-food items.

The study produced models for predicting both potential *P. truncatus* infestation and the grain damage caused by the pest together with other grain storage pests in contrasting environments. The effect of seasonality on the *P. truncatus* infestation and grain damage was also shown to be a contributing factor to the state of the grain at different times during a typical southern African grain storage season. The high accuracy produced using the machine learning approach demonstrated the clear potential of solving real agricultural issues by starting with the small datasets that exist and refining the models based on new input data and structured collection of such data. A larger dataset can produce models with better accuracy over a wider spectrum of observations and conditions. The models developed can packaged to aid extension staff in advising farmers on the timing of grain treatment based on the state of grain expected if no treatment is applied after harvest.

The study presented an alternative approach to working with data from field and exploratory experiments using machine learning and open source software packages. The models developed present a methodology to iteratively improve prediction of natural processes from research. The models can be used to create decision-support tools that can run on various platforms such as mobile applications which are increasingly becoming available to farmers and stakeholders as low-cost information gathering and dissemination devices.

Acknowledgements

The authors would like to thank the European Union (EU) for funding the research project "*Supporting smallholder farmers in southern Africa to better manage climate-related risks to crop production and postharvest handling*" through the Food and Agriculture Organization (FAO) (Grant No. DCI-FOOD/2012/304-807). Our gratitude is also extended to the trial host farmers and national extension services for their cooperation, contributions and active participation.

References

Abernethy, M. 2010. Data mining with WEKA, Part 2: Classification and clustering. *IBM developer Works*. Available: http://www.ibm.com/developerworks/library/os-weka2/os-weka2-pdf.pdf [2019, February 15].

APHLIS. 2018. Zimbabwe cereal postharvest losses. Available: http://www.erails.net/ZW/aphlis/aphlis-zimbabwe/ [2018, February 23].

Arthur, F.H., Morrison, W.R., Morey, A.C., 2019. Modeling the potential range expansion of larger grain borer, *Prostephanus truncatus* (Coleoptera: Bostrichidae). Scientific Reports 9, 6862. https://doi.org/10.1038/s41598-019-42974-5.

Barros, R.C., de Carvalho, A.C.P.L.F. & Freitas, A.A. 2015. *Automatic design of decision-tree algorithms*. (no. 9783319142302). DOI: 10.1007/978-3-319-14231-9_4.

Bhagawati, K., Sen, A., Shukla, K. kumar & Bhagawati, R. 2016. *Application and scope of data mining in agriculture*. V. 3.

Birkinshaw, L.A., Hodges, R.J., Addo, S. & Riwa, W. 2002. Can “bad” years for damage by *Prostephanus truncatus* be predicted? *Crop Protection*. 21(9):783–791. DOI: 10.1016/S0261-2194(02)00038-8.

Blockeel, H. & Struyf, J. 2001. Efficient algorithms for decision tree cross-validation. *Journal of Machine Learning Research*. 3(4–5):9. DOI: 10.1162/jmlr.2003.3.4-5.621.

Bonu-Ire, M., T Millicent Captain-Esoah, S. & Angyiereyiri, E.D. 2015. *Predation and parasitisation of* Prostephanus truncatus *by* Teretrius nigrescens *and* Anisopteromalus calandrae *respectively under controlled environmental conditions*. V. 5.

Borgemeister, C., Adda, C., Sétamou, M., Hell, K., Djomamou, B., Markham, R.H. & Cardwell, K.F. 1998. Timing of harvest in maize: Effects on post-harvest losses due to insects and fungi in central Benin, with particular reference to *Prostephanus truncatus* (Horn) (Coleoptera: Bostrichidae). *Agriculture, Ecosystems and Environment*. 69(3):233–242. DOI: 10.1016/S0167-8809(98)00109-1.

Boxall, R.A. 2002. Damage and loss caused by the larger grain borer *Prostephanus truncatus*. *Integrated Pest Management Reviews*. 7(2):105–121. DOI: 10.1023/A:1026397115946.

CAB International. 2018. *Prostephanus truncatus (larger grain borer)*. Available: https://www.cabi.org/isc/datasheet/44524 [2018, November 06].

Chigoverah, A.A. & Mvumi, B.M. 2018. Comparative efficacy of four hermetic bag brands against *Prostephanus truncatus* (Coleoptera: Bostrichidae) in stored maize grain. *Journal of Economic Entomology*. 111(5):2467–2475. DOI: 10.1093/jee/toy217.

Cunningham, S.J. & Holmes, G. 1999. Developing innovative applications in agriculture using data mining. In *SEARCC’99 conference proceedings*. 12. Available: https://www.cs.waikato.ac.nz/~ml/publications/1999/99SJC-GH-Innovative-apps.pdf.

Danho, M., Haubruge, E., Gaspar, C. & Lognay, G. 2000. *Selection of grain-hosts by Prostephanus truncatus (Coleoptera, Bostrychidae) in the presence of Sitophilus zeamais (Coleoptera, Curculionidae) previously infested grains*. V. 130.

Donatelli, M., Magarey, R.D., Bregaglio, S., Willocquet, L., Whish, J.P.M. & Savary, S. 2017. Modelling the impacts of pests and diseases on agricultural systems. *Agricultural Systems*. 155:213–224. DOI: 10.1016/j.agsy.2017.01.019.

Dunstan, W.R. & Magazini, I. 1980. Outbreaks and new records. Tanzania. The larger grain borer on stored products. *FAO Plant Protection Bulletin*. 29:80–81.

edX. 2018. *Visualizing Data with Python*. Available: /course/data-visualization-with-python [2019, June 21].

Fadamiro, H.Y. & Wyatt, T.D. 1995. Flight initiation by *Prostephanus truncatus* in relation to time of day, temperature, relative humidity and starvation. *Entomologia Experimentalis et Applicata*. 75(3):273–277. DOI: 10.1111/j.1570-7458.1995.tb01937.x.

Fadamiro, H.Y., Gudrups, I. & Hodges, R.J. 1998. Upwind flight of *Prostephanus truncatus* is mediated by aggregation pheromone but not food volatiles. *Journal of Stored Products Research*. 34(2–3):151–158. DOI: 10.1016/S0022-474X(97)00044-1.

Faith, J. 2007. Targeted projection pursuit for interactive exploration of high- dimensional data sets. In *11th International Conference Information Visualization - Supplements (IV '07)(IV)*. 286–292. DOI: 10.1109/IV.2007.107.

FAO Ed. 2018. *Building climate resilience for food security and nutrition*. (The state of food security and nutrition in the world no. 2018). Rome: FAO.

Frank, E., Wang, Y., Inglis, S., Holmes, G. & Witten, I.H. 1998. Using model trees for classification. *Machine Learning*. 32(1):63–76. DOI: 10.1023/A:1007421302149.

Fritz, H., Saïd, S., Renaud, P.-C., Mutake, S., Coid, C. & Monicat, F. 2003. The effects of agricultural fields and human settlements on the use of rivers by wildlife in the mid-Zambezi valley, Zimbabwe. *Landscape Ecology*. 18(3):293–302. DOI: 10.1023/A:1024411711670.

FutureLearn. 2019. *Evaluating clusters - more data mining with Weka*. Available: https://www.futurelearn.com/courses/more-data-mining-with-weka/0/steps/29129 [2019, March 14].

Giles, P.H. & Leon, O. 1974. Infestation problems in farm-stored maize in Nicaragua. In *Proceedings of the 1st International Working Conference on Stored Products Entomology*. Savannah, Georgia. 68–76.

Golob, P. & Hanks, C. 1990. Protection of farm stored maize against infestation by *Prostephanus truncatus* (Horn) and *Sitophilus* species in Tanzania. *Journal of Stored Products Research*. 26(4):187–198. DOI: 10.1016/0022-474X(90)90021-J.

Gonzalez-Sanchez, A., Frausto-Solis, J. & Ojeda-Bustamante, W. 2014. Predictive ability of machine learning methods for massive crop yield estimation. *Spanish Journal of Agricultural Research*. 12(2):57–65.

Grabczewski, K. 2014. *Meta-learning in decision tree induction. Studies in Computational Intelligence 498*. 1st ed. Springer.

Hall, M.A. 1999. Correlation-based feature subset selection for machine learning. PhD. New Zealand Department of Computer Science, Waikato University.

Hodges, R.J. 1986. The biology and control of *Prostephanus truncatus* (Horn) (Coleoptera: Bostrichidae)—A destructive storage pest with an increasing range. *Journal of Stored Products Research*. 22(1):1–14. DOI: 10.1016/0022-474X(86)90040-8.

Hodges, R.J. 2002. Detection and monitoring of larger grain borer, *Prostephanus truncatus* (Horn) (Coleoptera: Bostrichidae). *Integrated Pest Management Reviews*. 7:223–243. DOI: 10.1023/B:IPMR.0000040815.06804.c1.

Hodges, R., Dunstan, W.R., Magazini, I. & Golob, P. 1983. An outbreak of *Prostephanus truncatus* (Horn) (Coleoptera: Bostrichidae) in East Africa (Tabora, Tanzania). *Protection Ecology*. 5(2):183–194.

Hodges, R.J., Addo, S. & Birkinshaw, L. 2003. Can observation of climatic variables be used to predict the flight dispersal rates of *Prostephanus truncatus*? *Agricultural and Forest Entomology*. 5(2):123–135. DOI: 10.1046/j.1461-9563.2003.00170.x.

Holst, N., Meikle, W.G. & Markham, R.H. 2000. Grain injury models for *Prostephanus truncatus* (Coleoptera: Bostrichidae) and *Sitophilus zeamais* (Coleoptera: Curculionidae) in rural maize stores in West Africa. *Journal of Economic Entomology*. 93(4):1338–1346.

IPCC. 2014. *Climate Change 2014: Synthesis report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Intergovernmental Panel on Climate Change. Available: https://www.ipcc.ch/site/assets/uploads/2018/05/SYR_AR5_FINAL_full_wcover.pdf.

Karim, F., Karim, F. & Frihida, A. 2017. Monitoring system using web of things in precision agriculture. *Procedia Computer Science*. 110:402–409. DOI: 10.1016/j.procs.2017.06.083.

Kenneth, D. 1988. A review of insect infestation of maize in farm storage in Africa with special reference to the ecology and control of *Prostephanus truncatus*. *Overseas Development Natural Resources Bulletin*. 18.

Kotthoff, L., Thornton, C., Hoos, H.H., Hutter, F. & Leyton-Brown, K. 2017. Auto-WEKA 2.0: Automatic model selection and hyperparameter optimization in WEKA. *Journal of Machine Learning Research*. 18(25):1–5.

Krall, S. 1984. New threat to farm-level maize storage in West Africa: *Prostephanus truncatus* (Horn) (Coleoptera: Bostrichidae). *Tropical stored products information*. 50:26–31.

Kuhn, K. & Johnson, K. 2013. *Applied Predictive Modelling*. Springer.

Maiorano, A., Cerrani, I., Fumagalli, D. & Donatelli, M. 2014. New biological model to manage the impact of climate warming on maize corn borers. *Agronomy and Sustainable Development*. 34:609–621. DOI: 10.1007/s13593-013-0185-2.

Majumdar, J., Naraseeyappa, S. & Ankalaki, S. 2017. Analysis of agricultural data using data mining techniques: application of big data. *Journal of Big Data*. 4(20). DOI: 10.1186/s40537-017-0077-4.

Makundi, R.H., Swila, N.N., Misangu, R.N., Reuben, S.W.M., Mwatawala, M., Sikira, A., Kilonzo, B.S., Lyimo, H., *et al*. 2010. Dynamics of infestation and losses of stored maize due to the larger grain borer (*Prostephanus truncatus* Horn) and maize weevils (*Sitophilus zeamais* Motschulsky). *Archives of Phytopathology and Plant Protection*. 43(14):1346–1355. DOI: 10.1080/03235400802425804.

Mansour, Y. 1997. Pessimistic decision tree pruning based on tree size. In *14th International Conference on Machine Learning*. 195–201.

Mashavave, T., Mapfumo, P., Mtambanengwe, F., Chikowo, R., Gwandu, T., Nezomba, H. & Siziba, S. 2011. Factors influencing participation of smallholder farmers in knowledge sharing alliances around SOFECSA field-based learning centres. In *10th African Crop Science Conference Proceedings, Maputo, Mozambique*. 10-13 October 2011. 335–338. Available: https://www.cabdirect.org/cabdirect/abstract/20133200269 [2019, February 21].

McQueen, R.J., Garner, S.R., Nevill-Manning, C.G. & Witten, I.H. 1995. Applying machine learning to agricultural data. *Computers and Electronics in Agriculture*. 12(4):275–293. DOI: 10.1016/0168-1699(95)98601-9.

Meikle, W.G., Holst, N., Scholz, D. & Markham, R.H. 1998. Simulation Model of *Prostephanus truncatus* (Coleoptera: Bostrichidae) in Rural Maize Stores in the Republic of Benin. *Environmental Entomology*. 27(1):59–69. DOI: 10.1093/ee/27.1.59.

Mendis, A. 2019. Data Visualization in Python: Matplotlib vs Seaborn. Available: https://www.kdnuggets.com/2019/04/data-visualization-python-matplotlib-seaborn.html [2019, June 21].

Microsoft. 2016. Data science essentials in Python. 3–5.

Mlambo, S., Mvumi, B.M., Stathers, T., Mubayiwa, M. & Nyabako, T. 2017. Field efficacy of hermetic and other maize grain storage options under smallholder farmer management. *Crop Protection*. 98:198–210. DOI: 10.1016/j.cropro.2017.04.001.

Mlambo, S., Mvumi, B.M., Stathers, T., Mubayiwa, M. & Nyabako, T. 2018. Field efficacy and persistence of synthetic pesticidal dusts on stored maize grain under contrasting agro-climatic conditions. *Journal of Stored Products Research*. 76. DOI: 10.1016/j.jspr.2018.01.009.

Moses, J.A., Jayas, D.S. & Alagusundaram, K. 2015. Climate change and its implications on stored food grains. *Agricultural Research*. 4(1):21–30. DOI: 10.1007/s40003-015-0152-z.

Muatinte, B.L. & Van den Berg, J. 2019. Suitability of Wild Host Plants and Firewood as Hosts of *Prostephanus truncatus* (Coleoptera: Bostrichidae) in Mozambique. *Journal of Economic Entomology*. DOI: 10.1093/jee/toz042.

Muatinte, B. L., Kavallieratos, N. G., Boukouvala, M. C., García-Lara, S., L. Margarita López-Castillo, L. M. and Mvumi, B. M. 2019. The threat of the larger grain borer, *Prostephanus truncatus* (Coleoptera: Bostrichidae) and practical control options for the pest. *CAB Reviews* 14:1–25.

Muatinte, B.L., Van Den Berg, J. & Santos, L.A. 2014. *Prostephanus truncatus* in Africa : a review of biological trends and perspectives on future pest management Strategies. *African Crop Science Journal*. 22(3):237–256.

Mubayiwa, M., Mvumi, B.M., Stathers, T.E., Mlambo, S. & Nyabako, T. 2018. Blanket application rates for synthetic grain protectants across agro-climatic zones: Do they work? Evidence from field efficacy trials using sorghum grain. *Crop Protection*. 109. DOI: 10.1016/j.cropro.2018.01.016.

Munyuri, S.W. & Tabu, I.M. 2013. Resistance to the larger grain borer (*Prostephanus truncatus*) and yield performance in selected local maize landraces in Kenya. *International Journal of Agronomy and Agricultural Research*. 3(2):40–47.

Mvumi, B.M. & Stathers, T.E. 2014. Food security challenges in Sub-Saharan Africa: The potential contribution of postharvest skills, science and technology in closing the gap. In *Proceedings of the 11th International Working Conference on Stored-Product Protection*. Chiang Mai, Thailand. 33–43. DOI: 10.14455/DOA.res.2014.7.

Mvumi, B.M., Golob, P., Stathers, T.E. & Giga, D.P. 2003. Insect population dynamics and grain damage in small farm stores in Zimbabwe, with particular reference to Sitotroga cerealella (Oliver) (Lepidoptera: Gelechiidae). In *Proceedings of the 8th International Working Conference on Stored Product Protection*. York, UK: CABI Publishing, Walingford, UK. 151–168. Available: http://agris.fao.org/agris-search/search.do?recordID=GB2012101557 [2019, February 21].

Mvumi, B.M., Morris, M., Stathers, T.E. & Riwa, W. 2008. Doing things differently: Post-Harvest Innovation Learning Alliances in Tanzania and Zimbabwe. In *Innovation Africa: Enriching farmers' livelihoods*. London: Taylor & Francis. Chapter 12. Available: https://www.academia.edu/19836698/Doing_things_differently_Post-Harvest_Innovation_Learning_Alliances_in_Tanzania_and_Zimbabwe [2019, February 21].

Nang'ayo, F.L.O., Hill, M.G., Chandi, E.A., Chiro, C.T., Nzeve, D.N. & Obiero, J. 1993. The natural environment as a reservoir for the Larger Grain Borer *Prostephanus truncatus* (Horn) (Coleoptera: Bostrichidae) in Kenya. *African Crop Science Journal*. 1(1):39–47. DOI: 10.4314/acsj.v1i1.54745.

Nansen, C., Korie, S., Meikle, W.G. & Holst, N. 2001. Sensitivity of *Prostephanus truncatus* (Coleoptera: Bostrichidae) flight activity to environmental variables in Benin, West Africa. *Environmental Entomology*. 30(6):1135–1143. DOI: 10.1603/0046-225X-30.6.1135.

Nyabako, T. & Manzungu, E. 2012. An Assessment of the adaptability to climate change of commercially available maize varieties in Zimbabwe. *Environment and Natural Resources Research*. 2(1):32. DOI: 10.5539/enrr.v2n1p32.

O'Grady, M.J. & O'Hare, G.M.P. 2017. Modelling the smart farm. *Information Processing in Agriculture*. 4(3):179–187. DOI: 10.1016/j.inpa.2017.05.001.

Onyari, E.K. & Ilunga, F.M. 2013. Application of MLP neural network and M 5 P model tree in predicting streamflow : a case study of Luvuvhu catchment, South Africa. *International Journal of Innovation, Management and Technology*. 4(1):11–15. DOI: 10.7763/IJIMT.2013.V4.347.

Patel, H. & Patel, D. 2016. Survey of android apps for agriculture sector. *International Journal of Information Sciences and Techniques*. 6(1/2):61–67. DOI: 10.5121/ijist.2016.6207.

Perez-Mendoza, J., Baker, J.E., Arthur, F.H. & Flinn, P.W. 1999. Effects of Protect-It on efficacy of *Anisopteromalus calandrae* (Hymenoptera: Pteromalidae) parasitizing rice weevils (Coleoptera: Curculionidae) in wheat. *Environmental Entomology*. 28(3):529–534.

Pham, X. & Stack, M. 2018. How data analytics is transforming agriculture. *Business Horizons*. 61(1):125–133. DOI: 10.1016/j.bushor.2017.09.011.

Quellhorst, H., Athanassiou, C.G., Bruce, A., Scully, E.D., Morrison, W.R., III, 2019. Temperature-mediated competition between the invasive larger grain borer (Coleoptera: Bostrichidae) and the cosmopolitan maize weevil (Coleoptera: Curculionidae). Environmental Entomology. https://doi.org/10.1093/ee/nvz151

Rashid, A. 2003. Global Information and Early Warning System on Food and Agriculture: appropriate technology and institutional development challenges. In *Early Warning Systems for Natural Disaster Reduction*. J. Zschau & A. Küppers, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg. 337–344. DOI: 10.1007/978-3-642-55903-7_42.

Rawlings, J.O., Pantula, S.G. & Dickey, D.A. 1998. *Applied regression analysis: a research tool*. 2nd ed. Springer-Verlag New York, Inc. DOI: 10.1007/b98890.

Régnière, J. 2009. Predicting insect continental distributions from species physiology. *Unasylva (English ed.)*. 60(231/232):37–42.

Rembold, F., Hodges, R., Bernard, M., Knipschild, H. & Léo, O. 2011. *The African Postharvest Losses Information System (APHLIS)*. (EUR - Scientific and Technical Research Reports). Luxembourg: Publications Office of the European Union. Available: http://publications.jrc.ec.europa.eu/repository/bitstream/JRC62618/lbna24712enc.pdf.

Sadiku, M.N.O., Kotteti, C.M.M. & Musa, S.M. 2018. Machine learning in agriculture. *International Journal of Advanced Research in Computer Science and Software Engineering*. 8(6):26–28. DOI: 10.23956/ijarcsse.v8i6.713.

Samuel, A.L. 1959. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*. 3(3):210–229.

Sakka, M.K., Athanassiou, C.G., 2018. Competition of three stored-product bostrychids on different temperatures and commodities. *Journal of Stored Products Research* 79, 34–39. https://doi.org/10.1016/j.jspr.2018.07.002

Savidan, A. 2002. Tritrophic interactions in maize storage systems. University of Neuchatel.

Shalev-Shwartz, S. & Ben-David, S. 2013. *Understanding machine learning: From theory to algorithms*. V. 9781107057. DOI: 10.1017/CBO9781107298019.

Sharma, R., Kumar, S. & Maheshwari, R. 2015. Comparative analysis of classification techniques in data mining using different datasets. *International Journal of Computer Science and Mobile Computing (IJCSMC)*. 44(12):125–134. DOI: 10.17762/ijritcc2321-8169.150818.

Sharov, A.A. 1995. *Modeling insect dynamics*. Gummerus Printing, Jyvaskyla (Finland). Available: http://mmc2.geofisica.unam.mx/cursos/gest/Articulos/Geostatistics/Modelling%20Insect%20Dynamics.htm [2017, November 15].

Shires, S.W. 1979. Influence of temperature and humidity on survival, development period and adult sex ratio in *Prostephanus truncatus* (Horn) (Coleoptera, Bostrichidae). *Journal of Stored Products Research*. 15(1):5–10. DOI: 10.1016/0022-474X(79)90018-3.

Soetaert, K. & Herman, P.M.J. 2009. *A practical guide to ecological modelling: using R as a simulation platform*. Springer Netherlands. Available: https://www.springer.com/gp/book/9781402086236 [2019, February 21].

Stathers, T., Lamboll, R. & Mvumi, B.M. 2013. Postharvest agriculture in changing climates: its importance to African smallholder farmers. *Food Security*. 5(3):361–392. DOI: 10.1007/s12571-013-0262-z.

Stathers, T.E., Riwa, W., Mvumi, B.M., Mosha, R., Kitandu, L., Mngara, K., Kaoneka, B. & Morris, M. 2008. Do diatomaceous earths have potential as grain protectants for small-holder farmers in sub-Saharan Africa? The case of Tanzania. *Crop Protection*. 27(1):44–70. DOI: 10.1016/j.cropro.2007.04.020.

Teng, P.S. & Savary, S. 1992. Implementing the systems approach in pest management. *Agricultural Systems*. 40(1):237–264. DOI: 10.1016/0308-521X(92)90023-H.

Thornton, C., Hutter, F., Hoos, H.H. & Leyton-Brown, K. 2013. Auto-WEKA: combined selection and hyperparameter optimization of classification algorithms. In *19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 847–855. DOI: 10.1145/2487575.2487629.

Thornton, P.K., Ericksen, P.J., Herrero, M. & Challinor, A.J. 2014. Climate variability and vulnerability to climate change: a review. *Global Change Biology*. 20(11):3313–3328. DOI: 10.1111/gcb.12581.

Tonnang, H.E.Z., Hervé, B.D.B., Biber-Freudenberger, L., Salifu, D., Subramanian, S., Ngowi, V.B., Guimapi, R.Y.A., Anani, B., *et al*. 2017. Advances in crop insect modelling methods—

Towards a whole system approach. *Ecological Modelling*. 354(June):88–103. DOI: 10.1016/j.ecolmodel.2017.03.015.

Trevino, A. 2019. *Introduction to K-means Clustering*. Available: https://www.datascience.com/blog/k-means-clustering [2019, March 14].

Vincent, V., Thomas, R.G. & Staples, R.R. 1960. *An agricultural survey of Southern Rhodesia. Part 1. Agro-ecological survey*. (no. 19621701163). Salisbury (S. Rhodesia): Government Printers. Available: https://www.cabdirect.org/cabdirect/abstract/19621701163 [2019, February 21].

Wahbeh, A.H., Al-Radaideh, Q.A., Al-Kabi, M.N. & Al-Shawakfa, E.M. 2011. A comparison study between data mining tools over some classification methods. *International Journal of Advanced Computer Science and Applications*. 1(3). DOI: 10.14569/SpecialIssue.2011.010304.

Wang, Y. & Witten, I.H. 1996. *Induction of model trees for predicting continuous classes*. (Working Paper). Available: https://researchcommons.waikato.ac.nz/handle/10289/1183 [2019, February 21].

Wang, D., Chen, T. & Dong, J. 2014. Research of the early warning analysis of crop diseases and insect pests. In *Computer and Computing Technologies in Agriculture VII*. D. Li & Y. Chen, Eds. (IFIP Advances in Information and Communication Technology). Springer Berlin Heidelberg. 177–187. Available: http://reliefweb.int/sites/reliefweb.int/files/resources/wfp274160.pdf.

Winner, W.A. 1985. Introduction to kernel density estimation. (December, 1):90.

Witten, I.H., Frank, E., Hall, M.A. & Pal, C.J. 2016. *Data mining, Fourth Edition: practical machine learning tools and techniques*. 4th ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Worner, J. 1998. Some problems and approaches to modelling insect phenology. In: J. Baumgärtner, P. Brandmayr, and B.F.J. Manly (eds.), Important aspects in population ecology. In *Proceedings of the 20th International Congress on Entomology*. Florence, Italy: A.A. Balkema, The Netherlands. 89–98.