BIOSYSTEMS ENGINEERING 166 (2018) 76–89

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Development of a linear mixed model to predict the picking time in strawberry harvesting processes

Farangis Khosro Anjom, Stavros G. Vougioukas, David C. Slaughter*

University of California, Davis, Department of Biological and Agricultural Engineering, One Shields Avenue, Davis, CA, 95616-5270, USA

ARTICLE INFO

Article history:
Received 24 March 2017
Received in revised form
1 September 2017
Accepted 18 October 2017

Keywords:
Statistics
Prediction
Specialty crops
Harvest aids
Manual harvesting
Time studies

In manual fruit and vegetable harvesting, picking time statistics can be used to improve labour management and optimise the design and operation of harvest-aiding machines, such as conventional cross-row conveyors or recently proposed robotic transport carts. In this study, a dataset of 161 picking times from 18 workers was collected in commercial strawberry fields in Salinas, California, and a set of conditional linear mixed models (LMMs) was formulated to model the amount of time ("picking time") required by a picker to fill an empty tray with harvested crop. The LMMs were based on different combinations of the following influencing factors: picker speed, time of day, plant spacing, and picking cart style. The significance of effects of these factors was investigated and the LMMs were compared with each other using cross-validation (CV) techniques. The LMMs were also evaluated using a new dataset collected during the next year's harvest season. The best predictive LMM was found to be a heterogeneous model with "picker speed", "time of day", and "picking cart" factors. The model had a prediction error of 134.9 s based on 10-fold CV, and 136.8 s based on leave-one-out CV (LOOCV). The selected model predicts *a priori* mean and standard deviation of picking times for any given combination of factor levels. For instance, if picker speed is 'fast', the time of day is 'morning', and the picking cart is 'standard', the marginal predicted picking time is 477.1 ± 42.4 s. The proposed methodology and model structures offer a practical tool for strawberry picking time modelling, which could also be applied to other manually harvested specialty crops such as raspberries, cherry tomatoes, and table grapes.

1. Introduction

Manual harvesting represents up to 60% of variable production costs for fresh market strawberries in central coast, California (Bolda, Tourte, Klonsky, & De Moura, 2010; Bolda, Tourte, Klonsky, De Moura, & Tumber, 2014). Additionally, increasing worker shortages can lead to loss of production. Mechanised labour aids aim to increase worker productivity by making it easier to pick or by reducing the non-productive parts of the harvesting cycle (Baugher et al., 2009). In California, for crops like strawberries and most fresh vegetables such as celery, broccoli, and cauliflower, long cross-row conveyors are being used to help reduce pickers' walking times, during delivery of their harvested crop. For cost efficiency purposes, such labour aids are long and extend over many rows, in order

* *Corresponding author.*
E-mail address: svougioukas@ucdavis.edu (S.G. Vougioukas).
https://doi.org/10.1016/j.biosystemseng.2017.10.006

Notation

10FCV	10-fold cross-validation
AIC	Akaike information criteria
ANOVA	analysis of variance
cAIC	conditional Akaike information criteria
CV	cross-validation
$CV_{(K)}$	root-mean-square error of KFCV (s)
F_k	picking time observations in fold k
G	covariance matrix of random effects
GLS	generalised linear squares
GPS	global positioning system
ID	picker ID
IRB	institutional review board
K	number of folds in KFCV
KFCV	K-fold cross-validation
LMM	linear mixed model
LOOCV	leave-one-out cross-validation
LOOCV	root-mean-square error of LOOCV (s)
MAPE	mean-absolute-percentage error
MSE_k	mean-square error in fold k (s^2)
n	total number of picking time observations
n_k	number of picking time observations in fold k
p	total number of fixed-effects parameters
PDF	probability density function
PRESS	predicted residual error sum of squares (s^2)
q	total number of random effects
R	covariance matrix of random errors
RBD	randomised block design
REML	restricted maximum likelihood
RMSE	root-mean-square error
r_c	conditional residuals (s)
SE	standard error
$s\{\hat{\beta}\}$	standard error of estimated β
$s\{\hat{y}_h\}$	standard error of marginal predicted mean picking time for a new combination of factor levels
t_e	time instant when a picker starts filling a tray
t_f	time instant when a picker finishes filling a tray
V	variance of Y
WLS	weighted least squares
X	$n\times p$ design matrix of fixed effects
X_h	$1\times p$ vector of indicator variables for a new combination of factor levels
Y	$n\times 1$ vector of observed picking time (s)
Z	$n\times q$ design matrix of random effects
β	$p\times 1$ vector of fixed-effects parameters
$\hat{\beta}$	estimated β
γ	$q\times 1$ vector of random effects
$\hat{\gamma}$	estimated γ
ε	$n\times 1$ vector of random error
Δt_{ef}	tray fill-up time or picking time (s)
Δt_{fe}	time spent between t_f and t_e (s)
$\widehat{\Delta t}_{ef}$	predicted Δt_{ef} (s)
$\hat{t}_f$	predicted t_f
$\hat{y}_h$	marginal predicted mean picking time for a new combination of factor levels (s)
$\hat{y}_i\|\gamma$	conditional predicted mean picking time given random effects (s)

to accommodate large crews (5–9 rows are typical). However, large multi-row machines are associated with higher acquisition and operating costs, problematic transportation and field deployment (e.g., time consuming turning at headlands), and potential worker underutilisation, since slow pickers may limit the machine's pace. As an alternative, teams of small harvest-aiding robots are being developed, which supply pickers with empty trays and transport full trays to unloading stations at the edges of a field (Vougioukas, Spanomitros, & Slaughter, 2012). A similar system is under development for automated bin transportation during orchard harvesting operations (Ye, He, & Zhang, 2016). Such distributed transportation systems can offer scalability, redundancy, and easier deployment.

In this study, the amount of time required by a picker to fill an empty tray with the harvested strawberries is referred to as the *picking time*; its inverse is the *picking throughput*. Picking time/throughput statistics can be used to improve labour management and optimise the design and operation of harvest-aiding machines. For example, the overall throughput of a multi-row machine is determined by factors such as yield, crew size, picker throughputs, and machine advancement speed. In the case of teams of robotic tray transporters, scheduling depends on the arrival rates of new transport requests, i.e., picker throughput. In dynamic pickup and delivery systems, it has been shown that prediction of the spatiotemporal distribution of future requests reduces customers' waiting times (Ichoua, Gendreau, & Potvin, 2006). Hence, prediction of each picker's next tray-transport request, i.e., when and where each picker's tray will be ready (full) for transportation, is essential for optimised robot scheduling, and consequently increased picking productivity.

Despite their economic importance, picking time models that can be used for design, optimisation, and operation of harvest aids for strawberries are not available. The main contribution of this paper lies in formulating and validating stochastic models that predict the expected value and standard deviation of picking time for manual strawberry harvesting. The proposed methodology and model structures could be applied to other manually harvested crops such as raspberries, cherry tomatoes, and table grapes. The models are stochastic because picking times depend on the complex interaction of several stochastic biological factors such as individual plant yield and its spatial distribution, individual workers' picking skills, and fatigue.

Time and motion studies in agricultural operations have been conducted in the past, mainly for improved labour management (e.g., Berlage & Langmo, 1982; Luxhoj & Giacomelli, 1990). Some results from time studies of manual harvesting have also been reported. The mean and standard deviation of picking times have been measured and used in simulations, in order to improve tomato-harvesting