Scientia Horticulturae 138 (2012) 165–170

Contents lists available at SciVerse ScienceDirect

Scientia Horticulturae

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

Prediction of harvest start date in highbush blueberry using time series regression models with correlated errors

Carlos Muñoz[a], Julio Ávila[a], Sonia Salvo[b,*], Juan I. Huircán[a]

[a] Department of Electrical Engineering, Universidad de La Frontera, Chile
[b] Department of Mathematics and Statistics, Universidad de La Frontera, Chile

A R T I C L E I N F O

Article history:
Received 8 January 2012
Accepted 16 February 2012

Keywords:
Blueberry
Vaccinium
Heat units
Time series regression
Harvest date

A B S T R A C T

This work proposes a model for predicting the harvest start date sufficiently far ahead to enable the farmer to make a well-informed plan. The heat unit method is widely used in agriculture as the phenological unit of time, which offers the least variation in date predictions, and heat units have been used to estimate the start of harvesting in various crops. The problem is that the farmer needs to know the number of days and not the number of heat units that are needed until the harvest can begin. It is proposed that the daily maximum and minimum temperature time series be modelled through regression models with errors correlated using a sine curve. Using the requirements reported by Carlson and Hancock (1991) for the start of harvest of 13 varieties of blueberry over 15 years, a model has been developed that allows the requirements of heat units to be translated into days remaining until harvest. The models are estimated at intervals of 3 months, 2 months, 1 month, 14 days and 7 days before the date at which the heat unit requirements are reached. Three months ahead, the error was less than 10 days late, and 7 days ahead, it was 2 days late. A blueberry orchard in Temuco, Chile, was used as a case study and had similar results. All the errors are within the variability of the heat unit models. The models can be used by farmers to predict and plan the blueberry harvest with adjustments for location and variety.

1. Introduction

Blueberry cultivation is an important economic activity in Chile, where it is practiced between the Valparaíso and Los Lagos regions and covers more than 7000 ha (ODEPA and CIREN, 2010). In the 2010 season, more than 50,000 tons of blueberries were exported, mainly to the United States and the European Union (Bravo, 2011). Blueberries can be exported either frozen or fresh; however, the latter draw a higher sale price in the final markets. In 2011, for example, frozen blueberries drew only 40% of the value of fresh fruit (ODEPA, 2011). These exports to the northern hemisphere are favoured by the high sale prices achieved in the early and late varieties produced in the off-season. There are two important aspects for fresh blueberry exporters to consider: estimating the yield, which has been addressed (Hancock et al., 2000; Salvo et al., 2012), and the harvest start date. Knowing the harvest start date accurately at least 2–3 weeks in advance allows coordination of the procedures required for marketing large volumes (Mainland, 2000). Moreover, with this coordination, pre-packing delays can be reduced and the cold chain needed to export fresh blueberries at an acceptable quality can be maintained (Jackson et al., 1999). The procedures to be coordinated are obtaining certified clamshells, hiring a sufficient number of qualified pickers, training new personnel, preparing the packing process and obtaining refrigerated transport.

The traditional method of estimating the harvest start date is to count the days from flowering; but this approach is subject to too much variability between seasons (Baptista et al., 2006). The variability in early varieties to reach 50% mature fruit is from 4 to 9 days between seasons (Lyrene and Sherman, 1984), with a variation coefficient between 6.5 and 8% (Gupton et al., 1996). For example, Mainland (2000) determined that the number of days elapsed from flowering to harvest may be between 52 and 62. However, the harvest date is highly correlated with the days needed for the fruits to reach maturity ($r = 0.718$) and is negatively correlated with the weight of the individual fruits ($r = -0.660$) (Suzuki and Kawata, 2001). This correlation is due to the blueberry phenology being highly dependent on climatic conditions and the development stages of the fruit. However, the accumulation of heat units is a more robust phenological indicator, which starts to accumulate from the end of the latency period (Carlson and Hancock, 1991). The heat unit method has been used for numerous crops, including soft fruit (Everaarts, 1999), sweet potato (Villordon et al., 2009), corn (Lass et al., 1993), brassica (Adak and Chakravarty, 2010), sugar cane (de Souza et al., 2011) and opium poppy (Kamkar et al., 2012). Heat

* Corresponding author at: Department of Mathematics and Statistics, Universidad de La Frontera, Av. Francisco Salazar 01145, Casilla 54-D, Temuco, Chile. Tel.: +56 45 325 000; fax: +56 45 592 822.
E-mail address: ssalvo@ufro.cl (S. Salvo).

0304-4238/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.scienta.2012.02.023

units are also used for predicting the start of harvests for fruit such as apple (Perry et al., 1987), cucumber (Perry and Wehner, 1990, 1996), banana (Umber et al., 2011), loquat (Hueso et al., 2007), musk melon (Jenni et al., 1998) and yellow pitaya (Nerd and Mizrahi, 1998). In rabbiteye blueberries, using heat units reduces the coefficient of variation associated with the days for fruit development, thus improving the prediction of the harvest start date (Carlson and Hancock, 1991; NeSmith, 2006).

As the accumulation of heat units depends on the maximum and minimum daily temperatures occurring during the year, the relation between the days elapsed and the heat accumulated does not remain constant from 1 year to another. The problem is that the farmer needs to know the number of days and not the number of heat units that are needed until the harvest can begin.

Hean and Cacho (2003) approximated the annual temperature fluctuation with a sine function, an approach that allows the accumulated heat to be converted to the days remaining to the start of the harvest. The objective of this work is, therefore, to predict the date on which the heat unit requirements are met and harvesting can begin. We also seek to determine the error, in days, when the prediction is made 3 months, 2 months, 1 month, 14 days and 7 days ahead to generate a degree of confidence that will enable the farmer to plan harvest logistics well in advance.

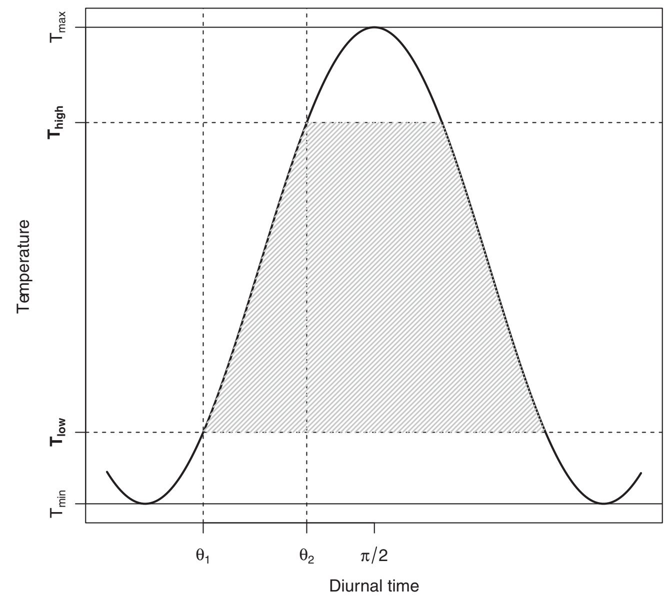

Fig. 1. Calculation of heat units per day as the area under the curve between the low and high thresholds of the daily temperature using a sine function.

2. Materials and methods

2.1. Heat model

In this study, the heat unit requirements for the start of blueberry harvesting reported by Carlson and Hancock (1991) are used. In that study, the authors investigated variability in the start of harvesting by comparing the heat unit model (heat units for the start of harvest) with the calendar day model (average date of the start of harvest) and concluded that there is less variability in the heat unit model. The study was carried out for 13 varieties of blueberry over 15 consecutive years in Bloomingdale, MI. For the heat unit model, the heat units were calculated using the Baskerville–Emin method, which uses high and low temperature thresholds (Baskerville and Emin, 1969). This method uses a sine curve approach to daily temperature based on the maximum ($T_{\max}$) and minimum ($T_{\min}$) daily temperatures. The daily quantity of heat is calculated as the area under the curve of this sine function between a low (T_{low}) and high (T_{high}) threshold (see Fig. 1).

The date for the start of heat unit accumulation and the low and high temperature thresholds for accumulation were determined using values that would reduce the variability in days to the start of harvesting. Table 1 shows the parameters determined for the heat unit and calendar day models. The parameters in the heat unit model were the start date of heat accumulation (SDATE), the low temperature threshold (T_{low}), and the high temperature threshold (T_{high}). These parameters were used to calculate the heat units to the start of harvesting (HU_m) and the variation in days to the start of harvesting once the heat accumulation was complete (σ_{hu}). For the calendar day model, the parameter used for the start of harvesting (HDATE) was the average of the last 15 years, which allowed the variation in days of HDATE (σ_{cal}) and the relation between the variations in days in the two models ($\sigma_{\text{hu}}/\sigma_{\text{cal}}$) to be obtained.

2.2. Databases

The daily temperature series (maximum, minimum and mean) recorded from 1974 to 1988, which was the period of Carlson and Hancock's study (1991), were obtained by the Bloomingdale, MI, meteorological station (42.38N, 85.96W, elevation 220.98 m). During these 15 years, the minimum temperature ranged from −30 °C to 24 °C, and the maximum ranged from −15 °C to 38 °C.

2.3. Time series model

The proposed models correspond to two linear regression models with errors correlated in discrete time, one for minimum temperature ($i = T_{\min}$) and the other for maximum temperature ($i = T_{\max}$).

$$T_i(k) = \beta_{i,0} + \beta_{i,1}\cos(\omega k) + \beta_{i,2}\sin(\omega k) + e_i(k), \quad i = T_{\min}, \quad T_{\max} \tag{1}$$

where $\beta_{i,j}$ are the parameters of the two models, k is a correlate that corresponds to the number of calendar days since 1 January 1988 and $e_i(k)$ is the error after elimination of the series trend in each of the models. The structure of both models includes a sine curve approximation, with a period of $\omega = 2\pi/365.2422$, to eliminate the annual trend of the series. The $\beta_{i,j}$ parameters are calculated by minimising the squared error between the maximum and minimum temperatures measured and those determined by these models.

The $e_i(k)$ errors were analysed to look for an autocorrelation greater than zero with the Durbin–Watson test. In the resulting ARMA(p,q) (Auto Regressive Moving Average) model, p is the order of the autoregressive model and q is the order of the moving average, given by Eq. (2).

$$\begin{aligned} e_i(k) &= \varphi_{i,1}e_i(k-1) + \varphi_{i,2}e_i(k-2) + \cdots + \varphi_{i,p}e_i(k-p) \\ &\quad + \theta_{i,1}\varepsilon_i(k-1) + \theta_{i,2}\varepsilon_i(k-2) + \cdots + \theta_{i,q}\varepsilon_i(k-q) \\ &\quad + \varepsilon_i(k) \quad i = T_{\min}, T_{\max} \end{aligned} \tag{2}$$

where $\varphi_{i,j}$ are the parameters of the autoregressive model, $\theta_{i,j}$ are the parameters of the moving average model and $\varepsilon_i(k)$ is the normal distribution error with an average of zero and a variance σ_i^2. The order of parameters p and q is obtained from the model with the smallest Akaike Information Criterion (AIC) (Akaike, 1974). The models were validated using tests on the residuals to verify the normality, independence and homogeneity of the variance.

For the 1988 season, the date when the heat unit requirements were met for the start of harvesting is called HUDATE. This date