Survival Analysis of Flower and Fruit Abortion in Sweet Pepper

A.M. Wubs and E. Heuvelink
Horticultural Production Chains Group
Department of Plant Sciences
Wageningen University
Marijkeweg 22, 6709 PG Wageningen
The Netherlands

L.F.M. Marcelis
Wageningen UR Greenhouse Horticulture
Wageningen University
P.O. Box 16, 6700 AA Wageningen
The Netherlands

L. Hemerik
Biometris
Department of Plant Sciences
Wageningen University
Bornsesteeg 47, 6708 PD Wageningen
The Netherlands

Keywords: *Capsicum annuum*, Cox proportional hazards model, source/sink ratio, source strength, plant density, inter-plant variation

Abstract

In order to obtain a crop growth model that can simulate inter- and intra-plant variation in fruit set, fruit abortion times in sweet pepper were analysed by means of survival analysis. Survival analysis is a statistical technique dealing with the timing of events. The Cox proportional hazards model estimates the instantaneous baseline probability of abortion per time (hazard rate), given that the fruit has not aborted yet. It also estimates the effect of explanatory factors (covariates) on the abortion rate. Two important factors known to influence abortion in sweet pepper are the supply and demand for assimilates (source and sink strength, respectively). A plant density experiment was analysed, as density influences plant source and indirectly also sink strength. Flowering as well as abortion or harvest dates were recorded for all individual flowers and fruits. LAI was measured to calculate the source strength with a photosynthesis-based simulation model. Empirical curves showed that survival of flowers and fruits (per plant) was higher at 1.6 plants/m^2 than at 3.1 or 4.6 plants/m^2. Plant density, source strength and source/sink ratio all correlated with the treatment plant density, but source/sink ratio explained most of the variation in the data in a Cox proportional hazards model. Averaging the source/sink ratio over seven days gave the best fit. Adding the position within the node (main or side branch) improved the fit; flowers from the side branch had a higher probability of abortion per unit time than flowers from the main branch. The susceptibility for fruit abortion differed among plants. The baseline hazard rate indicated that between 0 and 13 days after anthesis, flowers were most susceptible for abortion.

INTRODUCTION

In general, process-based crop growth models simulate the growth of average plants. However, in reality a lot of variation exists in growth and development between plants and between similar organs within one plant. A crop growth model taking this variation into account, for example by using a stochastic function, can predict variations in crop growth more realistically. Because the output of such a model is different for every model run, confidence intervals for expected yield or crop growth can be obtained. One of the possibilities to include this variation is by implementing results of a survival analysis into a crop growth model.

Survival analysis is a statistical technique for the analysis of timing of events (Kleinbaum, 1995). An advantage of survival analysis is that it can cope with so-called censored data. In censored data, the event of interest did not occur before the end of the experiment. The information that the event did not happen for a particular time is used in the analysis. Cox proportional hazards model is a commonly used model in survival analysis. With this model, a function is obtained which gives the probability per time unit of an event happening, given that it has not occurred yet. This is called the hazard rate. It is calculated from a baseline hazard rate, where all explanatory factors (covariates) are assumed to be zero, which is modified for the effect of the covariates, to account for factors increasing or decreasing the baseline hazard rate (see below).

The event of interest in this paper is abortion of sweet pepper flowers and fruits. Under given conditions, high variation among plants exists with respect to patterns of fruit abortion in sweet pepper plants (Heuvelink and Körner, 2001). Factors influencing fruit abortion are quite well known; the amount of available assimilates (source strength) and the demand for assimilates (sink strength) play an important role (Heuvelink et al., 2004; Marcelis et al., 2004), but also hormones have some influence (Huberman et al., 1997). In short, higher source strength decreases the percentage of abortion while higher plant sink strength increases the percentage of abortion. However, quantification of these effects is difficult and therefore fruit set is one of the weakest parts in crop growth models.

In this paper, an experiment in which the source strength of the plants was varied is analysed with survival analysis. This reveals the factor best explaining abortion of the fruits of sweet pepper and how it affects the timing of abortion.

MATERIALS AND METHODS

Experiments

Sweet pepper *Capsicum annuum* 'Mazurka' was grown in a Venlo-type greenhouse in Wageningen (The Netherlands) from January to July at three different plant densities: 4.6, 3.1 and 1.6 plants/m^2. Plant density influences the amount of assimilates available per plant (source strength). The experiment was set up in a randomized complete block design, with 7 blocks and 3 plots per block. The low and high density had 8 plants per block, the intermediate density 12 plants. Plants were grown in Rockwool slabs and temperature was set at 20°C. The two first order branches of each plant were retained. Of each dichotomic branching, the weaker branch was pruned just above the first leaf (= side branch), while the stronger one was allowed to grow (= main branch).

In each block, 1 plant per plot was used for observations on flowering, fruit abortion and fruit harvest. Every five days, measurements on length and diameter of the growing fruits on these plants were done to keep track of fruit growth. Six times during the experiment, one plant from each plot was harvested destructively. Of these plants, the number of leaves, leaf area and dry weight of leaves, stems and fruits and the position of fruits were registered. Ripe fruits (red) from all plants were harvested weekly.

Calculating Source and Sink Strength

Leaf area data from the destructive measurements were used to simulate growth of the sweet pepper plants using the Hortisim model (Gijzen et al., 1998), adapted for sweet pepper (Marcelis et al., 2006). In this way, the amount of assimilates available for plant growth (source strength) was obtained. For the three final sample dates in the high density, the simulation output (source strength) underestimated the measured dry matter production. For this case, the simulation output was adjusted to the measured total dry matter production by a scaling factor. The sink strength of each fruit, defined by its potential growth rate, was calculated using the first derivative of the Richards function, analogous to what Marcelis (1994) did for cucumber. Sink strength of a fruit is a function of its accumulated temperature sum after anthesis. Estimates for coefficients were obtained by fitting the data of Marcelis and Baan Hofman-Eijer (1995) to the Richards function. Total plant sink strength was calculated as the sum of all individual fruit sink strengths and the vegetative sink strength. Vegetative sink strength was assumed to depend only on temperature, i.e. higher with higher temperature (Marcelis et al., 2006).

Survival Analysis

The event of interest was abortion of flowers and fruits. No distinction was made between abortion of flowers and of fruits. The survival time started at flowering and ended when abscission of the flower or fruit was observed. An observation was censored when the fruit was harvested or when the flower or fruit was still on the plant by the end of the experiment. In both cases, the event of interest, abscission, had not (yet) occurred. In the former, the time from anthesis till harvest was the survival time, in the latter case, the last day of the experiment was taken as the end of the censored survival time. If a fruit was present on the plant, but did not show any growth for more than five days, the fruit was assumed to have aborted after the day when last growth was measured.

First, the resulting survival times were analysed using the Kaplan-Meijer technique (Kleinbaum, 1995). This yields the empirical survival curves for the three densities. The difference in survival between the densities was tested with the log-rank test. A disadvantage of this method is that the effects of covariates cannot be quantified.

Next, the data were analysed using the Cox proportional hazards model (Kleinbaum, 1995) to be able to quantify the effects of covariates. This method can handle several covariates and can also deal with time-dependent covariates, where the value can vary daily, for example source strength and source/sink ratio. Fixed covariates are covariates with a constant value during the survival time of the individual, like plant density or position on the plant. Cox proportional hazards model works with a hazard rate (equation 1), in this case the probability of abortion per time.

$$h(t,X) = h_0(t) * \exp[\sum_{i=1}^{n} \beta_i X_i(t)] \qquad \text{equation 1}$$

Here, $h(t,X)$ is the probability of abortion per time (hazard rate) at time t (days), influenced by fixed or time-dependent covariates $X_i(t)$ (i=1,..,n): $h_0(t)$ is the baseline probability of abortion at time t, which is multiplied by the exponential expression, where $X_i(t)$ are the values of the covariates at time t and β_i the coefficient quantifying the effect of the covariate X_i on the probability of abortion. If X_i is a fixed covariate, $X_i(t)$ is constant.

To allow for inter-plant differences, a frailty term was added. Fruits on a plant with a higher frailty are more susceptible for abortion. Equation 1 is extended with the extra term $\exp(\omega_j)$, where ω_j is the frailty of plant j (equation 2).

$$h(t,X) = h_0(t) * \exp(\omega_j) * \exp[\sum_{i=1}^{n} \beta_i X_i(t)] \qquad \text{equation 2}$$

$\text{Exp}(\omega_j)$ has a gamma distribution with mean one and unknown variance θ. The variance is estimated together with the baseline hazard and coefficients of the covariates. The degrees of freedom associated with the frailty are not preset (Therneau and Grambsch, 2000).

In the Cox proportional hazards model, the baseline hazard $h_0(t)$ has no predefined shape. It is estimated simultaneously with the regression coefficients β_i (i = 1,..,n) for the covariates and the frailty. The estimation is done by maximizing the partial likelihood function based on the data (Kalbfleisch and Prentice, 2002).

Plant density, source strength and source/sink ratio can be seen as covariates influencing the abortion related to the treatment plant density. First, these covariates were evaluated separately in a Cox proportional hazards model to see which of these explained most of the variation in the data. The time-dependent covariates source strength and source/sink ratio were for every day averaged over 1, 5, 7 or 10 days before the current day (further called interval), because it is unlikely that only the value of the preceding day determines abortion. The best covariate and the best interval for the time-dependent covariates were chosen based on the log-likelihood. The likelihood of a data point is the probability that the data point occurs under the assumption that the model is true. The log-likelihood is the natural logarithm of the product of all likelihoods. The model with the maximum log-likelihood (closest to zero) was considered the best fit and selected. Next, the position within the node (main or side branch) and variation between plants (frailty) were added to the selected model.

All analyses were done in the R language version 2.2.1, available through the internet (www.r-project.org) under the general public license. Significant differences between two models were tested with the likelihood ratio test (LR ratio). It compares the log-likelihood ratio of the full and the reduced model: 2*(log-likelihood$_{full}$ - log-likelihood$_{reduced}$). The LR-ratio has a chi-square distribution with degrees of freedom equal to the difference between the number of parameters of the full and of the reduced model. The significance of the coefficients related to each covariates in a model is tested with the Wald test statistic $\frac{\beta_i^2}{\mathrm{var}(\beta_i)}$, which has a chi-square distribution with degrees of freedom equal to the total number of covariates in the model (Haccou and Hemerik, 1985).

RESULTS

Empirical Survival Curves (Kaplan Meijer Analysis)

Plants grown at low plant density (1.6 plants m^{-2}) had a significantly higher fraction of fruits that did not abort before the end of the observation period (survival) than plants grown at high or intermediate plant densities (4.6 and 3.1 plants m^{-2}, respectively, P<0.001, Fig. 1). The fruits on plants of the high and intermediate plant density did not differ in their survival (P=0.46). The survival does not decrease to zero, because not all flowers abort and their survival times are censored due to harvest or ending of the experiment. In all three plant densities, the period in which most flowers abort is the first 10 days after flowering.

Effect of Factors Influencing Abortion and Sensitive Period for Abortion (Cox Proportional Hazards Model)

Source/sink ratio could explain most of the variation in the survival data, followed by source strength and density. Source/sink ratio fitted best to the data when averaged over seven days (Table 1). The log-likelihood without covariates (assuming no model, only a baseline hazard rate is considered) was -8898.7, implying that all fits of Table 1 were better than assuming no model for abortion (P<0.001).

Fitting the model with the covariates source/sink ratio (averaged over seven days), position within the node and frailty gave a log-likelihood of -8782.4, which means that this model is better than the model with only source/sink ratio (P<0.001). The Wald-tests for source/sink ratio as well as position within node indicated significant contributions to the model (Table 2). There was no interaction between source/sink ratio and position within node (P=0.48). The frailty variance was rather small, 0.0293, but nonetheless highly significant (P<0.001). Source/sink ratio had a negative coefficient β (Table 2), meaning a decrease in the probability of abortion per day if source/sink ratio increased. The coefficients β quantify the effect of the covariate on the probability of abortion per day when it is increased by one unit, given that all other circumstances are the same. For example, the coefficient for the covariate position within a node is 0.19. This means that if this covariate is 1 (side branch) instead of 0 (main branch), the probability of abortion per day is multiplied by exp(0.19) = 1.21, which means the probability of abortion per day for a flower on the side branch is 21% higher than for a flower on the main branch.

The cumulative probability per day of abortion is given in Figure 2. The slope of the cumulative probability is thus the instantaneous probability of abortion; a steeper slope means a higher instantaneous probability. The most vulnerable period for the flowers and fruits was from flowering to 13 days after flowering (dashed line). Also between 13 and 20 days, the probability of abortion per time unit was quite high, but after 20 days, the probability to abort is very low.

DISCUSSION

This paper shows the first results of our attempts to analyse fruit abortion in sweet pepper by means of survival analysis. Up to now, there are only a few examples of survival analysis used in plant sciences (Dungan et al., 2003; Vermerris and McIntyre, 1999).

The analysis with Cox proportional hazards model shows that source/sink ratio plays an important role in fruit abortion, as was already shown by previous authors (Marcelis et al., 2004; Turner and Wien, 1994). If the source/sink ratio is higher, there are more assimilates to distribute to the individual flowers/fruits and so the probability of abortion decreases. Source/sink ratio is highly correlated with plant density and source strength, but varies more between the plants than the latter two. The source strength was calculated to be the same for each plant within a density, while the sink strength and hence source/sink ratio was calculated per plant. This caused more variation in source/sink ratio between plants and a better fit to the data. The position of the fruit on the plant also had impact on the probability of abortion per day; fruits on the side branch had a higher probability of abortion per day. When pruning the plant, the smaller/weaker branch is pruned to form the side branch, while the stronger one becomes the main branch. This may explain the higher failure rate for the side branch. From the buds that did not reach flowering, the majority was also located on the side branch. The significant frailty refers to higher sensitivity of certain plants for fruit abortion. This might be due to variation in growth conditions in the greenhouse. The vulnerable period for abortion derived from the hazard function is comparable to what is found in literature (Marcelis et al., 2004).

Plant sink strength was not used in the model because there were high correlations with source strength and its effect could not be estimated accurately. The experiment lasted from January till July. During this time, the source strength increased, resulting in a better fruit set, consequently an increased sink strength. It would have been better if the experiment lasted over summer to avoid this parallel increase of source and sink strength. Daily temperature might also be a covariate in a survival analysis of sweet pepper flowers and fruit, but was not used because it was not affected by the treatment plant density.

Regarding the present analysis, some points in the analysis need to be improved. The time to abscission of the fruit was analysed, while abortion may take place already some time before the actual abscission happens and induction of abortion even earlier. A lag-time might be needed to account for this. A direct temperature effect on abortion was not included in the analysis, but the indirect effect via sink and source strength was taken into account. The temperature was increasing slightly during the season, but did not reach extreme high or extreme low values (average daily temperatures ranged from 19.1 to 27.0°C). However, it is known that low temperatures (Pressman et al., 1998) as well as high temperatures (Erickson and Markhart, 2002) increase abortion, due to malformation of pollen and consequently reduced pollination and fruit set. To get a good quantification of the effect of temperature on abortion, an experiment in which plants are grown at different temperatures should be conducted. Further, only one cultivar was analysed. Turner and Wien (1994) showed that there are cultivar differences in susceptibility to abortion due to different light levels. More experiments are needed to see if this cultivar difference can be quantified in a parameter, or that separate baseline hazard rates need to be defined. Methodologically, the Cox proportional hazards model assumes proportionality of the hazard rates for fixed covariates, meaning that the ratio of hazard rates of two individuals is constant over time (Kleinbaum, 1995). This assumption will be tested and corrections for non-proportionality of the data might be necessary.

After correcting for the points mentioned above, the results will be implemented in a crop growth model. Due to the stochastic survival function, output of each simulation run will differ. In this way, the model is able to compute confidence intervals for the output variables. After verification experiments are conducted, the model will be used for scenario studies.

ACKNOWLEDGEMENTS

We thank Santiago Dogliotti and Ana Mantzouranaki for doing the experiment.

Literature Cited

Dungan, R.J., Duncan, R.P. and Whitehead, D. 2003. Investigating leaf life spans with interval-censored failure time analysis. New Phytologist 158:593-600.

Erickson, A.N. and Markhart, A.H. 2002. Flower developmental stage and organ sensitivity of bell pepper (*Capsicum annuum* L.) to elevated temperatures. Plant, Cell and Environ. 25:123-130.

Gijzen, H., Heuvelink, E., Challa, H., Dayan, E., Marcelis, L.F.M., Cohen, S. and Fuchs, M. 1998. Hortisim: a model for greenhouse crop and greenhouse climate. Acta Hort. 456:441-450.

Haccou, P. and Hemerik, L. 1985. The influence of larval dispersal in the cinnabar moth (*Tyria jacobaeae*) on predation by the red wood ant (*Formica polyyctena*): an analysis based on the proportional hazards model. J. Animal Ecol. 54:755-769.

Heuvelink, E. and Körner, O. 2001. Parthenocarpic fruit growth reduces yield fluctuation and blossom-end rot in sweet pepper. Ann. Bot. 88:69-74.

Heuvelink, E., Marcelis, L.F.M. and Körner, O. 2004. How to reduce yield fluctuations in sweet pepper? Acta Hort. 633:349-355.

Huberman, M., Riov, J., Aloni, B. and Goren, R. 1997. Role of ethylene biosynthesis and auxin content and transport in high temperature-induced abscission of pepper reproductive organs. J. Pl. Gr. Reg. 16:129-135.

Kalbfleisch, J.D. and Prentice, R.L. 2002. The statistical analysis of failure time data. Wiley series in probability and statistics. Wiley, New York.

Kleinbaum, D.G. 1995. Survival analysis, a self learning text. Springer-Verlag, New York.

Marcelis, L.F.M. 1994. A simulation model for dry matter partitioning in cucumber. Ann. Bot. 74:43-52.

Marcelis, L.F.M. and Baan Hofman-Eijer, L.R. 1995. Growth analysis of sweet pepper fruits (*Capsicum annuum*). Acta Hort. 412:470-478.

Marcelis, L.F.M., Heuvelink, E., Baan Hofman-Eijer, L.R., Den Bakker, J. and Xue, L.B. 2004. Flower and fruit abortion in sweet pepper in relation to source and sink strength. J. Exp. Bot. 55:2261-2268.

Marcelis, L.F.M., Elings, A., Bakker, M.J., Brajeul, E., Dieleman, J.A., de Visser, P.H.B. and Heuvelink, E. 2006. Modelling dry matter production and partitioning in sweet pepper. Acta Hort. 718:121-128.

Pressman, E., Moshkovitch, H., Rosenfeld, K., Shaked, R., Gamliel, B. and Aloni, B. 1998. Influence of low night temperatures on sweet pepper flower quality and the effect of repeated pollinations, with viable pollen, on fruit setting. J. Hort. Sci. Biotech. 73:131-136.

Therneau, T.M. and Grambsch, P.M. 2000. Modelling survival data: extending the Cox model. Springer, New York.

Turner, A.D. and Wien, H.C. 1994. Dry matter assimilation and partitioning in pepper cultivars differing in susceptibility to stress-induced bud and flower abscission. Ann. Bot. 73:617-622.

Vermerris, W. and McIntyre, L.M. 1999. Time to flowering in brown midrib mutants of maize: an alternative approach to the analysis of developmental traits. Heredity 83:171-178.

Tables

Table 1. The log-likelihood of the explanatory factors (covariates) related to plant density when used as single covariate in the Cox proportional hazards model. The time-dependent covariates source strength and source/sink ratio were averaged over 1, 5, 7 and 10 days before the present day (interval). Plants were grown at 1.6, 3.1 and 4.6 plants/m^2. The bold number indicates the best fitting model.

Covariate	Interval (days) for time-dependent covariates			
	1	5	7	10
Density	-8888	-	-	-
Source strength	-8850	-8839	-8840	-8844
Source/sink ratio	-8837	-8814	**-8810**	-8820

Table 2. The explanatory factors (covariates), their coefficients β in the Cox proportional hazard model, effect on the probability of abortion per time unit (hazard rate) and significance. Plants were grown at 1.6, 3.1 and 4.6 plants/m^2.

Covariate	β	exp(β)	P
Source/sink ratio	-0.53	0.59	<<0.001
Position within node*	0.19	1.21	0.003

* main (0) or side (1) branch

Figures

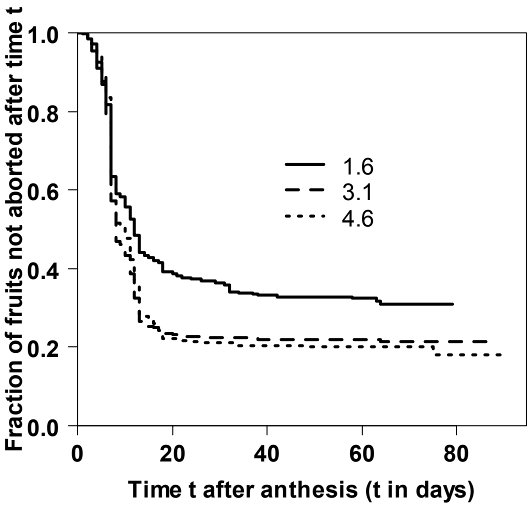

Fig. 1. The empirical survival curves of flowers in sweet pepper plants grown at 1.6, 3.1 and 4.6 plants/m^2 in a greenhouse from January to July. The graph shows for each day the fraction of flowers and fruits that has not aborted yet.

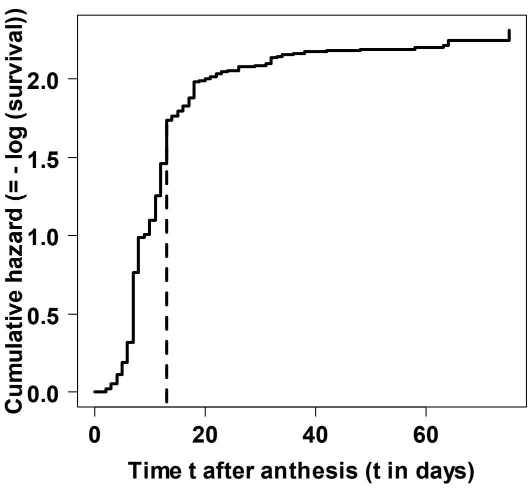

Fig. 2. The cumulative probability of abortion per day with all covariates set to zero (baseline hazard) for the abortion of flowers in sweet pepper, obtained from Cox proportional hazard analysis of a glasshouse experiment with three plant densities (1.6, 3.1 and 4.6 plants m^{-2}). Dotted line indicates end of period during which most flowers abort.