

Beta Regression Model for Predicting the Development of Pink Rot in Potato Tubers During Storage

S. KR. Yellareddygar, Julie S. Pasche, and Raymond J. Taylor, Department of Plant Pathology, Su Hua, Department of Statistics, and Neil C. Gudmestad, Department of Plant Pathology, North Dakota State University, Fargo 58105

Abstract

Yellareddygar, S. KR., Pasche, J. S., Taylor, R. J., Hua, S., and Gudmestad, N. C. 2016. Beta Regression model for predicting the development of pink rot in potato tubers during storage. *Plant Dis.* 100:1118-1124.

Pink rot is an important disease of potato with worldwide distribution. Severe yield and quality losses have been reported at harvest and in postharvest storage. Under conditions favoring disease development, pink rot severity can continue to increase from the field to storage and from storage to transit, causing further losses. Prediction of pink rot disease development in storage has great potential for growers to intervene at an earlier stage of disease development to minimize economic losses. Pink rot disease is estimated as percent rot confined on the interval (0 or 1, corresponding to 0% as no disease and 100% as maximum disease). In this study, beta regression is considered over the traditional ordinary least squares regression (linear regression) for fitting continuous response variables bounded on the unit interval (0,1). This method is considered a good alternative to data transformation and analysis by linear regression. The percentages of

incidence of pink rot in tubers at harvest, yield, and days after harvest were used as study covariates to predict pink rot development from 32 to 78 days postharvest. Results demonstrate that the interaction between percentage of pink rot at harvest and yield is a significant predictor ($P < 0.0001$) of the beta regression model. A linear regression model was also designed to compare the results with the proposed beta regression model. Linear predictors observed in diagnostic plots with linear regression model was found to not be constant and an adjusted R^2 (0.49) was obtained. The pseudo R^2 (0.56) and constant variance for this study suggests that the beta regression function is adequate for predicting the development of pink rot during storage. The use of the beta prediction model could help growers decide whether to apply a fungicide to tubers going into storage or to market their crop before significant storage losses are incurred.

Pink rot is an important soilborne disease of potato (*Solanum tuberosum*, L) caused by the oomycete *Phytophthora erythroseptica* Pethybr. (Boothroyd 1951; Cairns and Muskett 1933; Carrol and Sasser 1974; Goss 1949; Lambert and Salas 2001). It is most severe in crops planted in low-lying soils having poor water drainage and high moisture retention capacity (Lambert and Salas 2001; Taylor et al. 2006). Oospores present in the soil or plant debris serve as the primary inoculum, initial infection usually occurs belowground on roots, and subsequent infection may be found on stems, stolons, and tubers of potato (Lambert and Salas 2001; Peters and Sturz 2001). Typically, tuber infection occurs through the infected stolon; however, *P. erythroseptica* zoospores can also infect tubers through eyes and wounds (Johnson et al. 2004; Salas et al. 2003). Pink-rot-infected tubers are soft and rubbery but the periderm often remains intact. When infected, internal tuber tissue exposed to air for a few minutes will display a characteristic pink discoloration (Goss 1949; Taylor et al. 2004). In advanced cases, the lenticels may become wet and swollen and exude liquid. The periderm can readily slough off under these conditions, exposing the infected tissue to bacteria and other secondary rot organisms. Rot symptoms are most severe during prolonged wet and warm weather conditions during late crop development, and further infection is known to occur in storage (Al-Mughrabi et al. 2007; Salas et al. 2000; Secor and Gudmestad 1999; Taylor et al. 2004).

Potato production in the United States is affected by pink rot, and severe losses in yield and quality of tubers at harvest and in postharvest storage have been reported (Blodgett 1945; Gudmestad et al. 2007; Taylor et al. 2011). Proper management of the disease in the field during the growing season is needed to assure desirable yield and tuber quality at harvest. In turn, pink rot incidence at harvest will

have a direct and substantial impact upon disease development in storage. Because most potato cultivars are susceptible to pink rot (Peters and Sturz 2001; Salas et al. 2003; Taylor et al. 2008) strategies implemented to manage the disease in the field typically deal with cultural practices. These usually focus on conditions and practices that favor infection and disease progression such as avoiding excessive irrigation, planting in well-drained soils, timing vine kill to insure proper skin set at harvest, and modifying tuber handling procedures during harvest, transportation, and storage. The latter is extremely important because wounds offer an excellent infection court for *P. erythroseptica* (Salas et al. 2000). In many instances, implementation of these and other cultural practices can reduce pink rot incidence; however, fungicide applications are routinely used as supplemental control measures, particularly in fields with a history of severe losses due to pink rot.

The phenylamide fungicide mefenoxam (metalaxy1) has been very effective in controlling diseases caused by *Phytophthora* spp., and potato growers have relied upon this chemistry to manage both late blight and pink rot since it first became available nearly 40 years ago. Unfortunately, efficacy of mefenoxam has decreased since resistance to the fungicide developed and became pervasive throughout the North American population of the pathogen (Venkataramana et al. 2010; Taylor et al. 2002). Although mefenoxam can still be effective in areas where the *P. erythroseptica* population is composed of mefenoxam-sensitive strains (Peters et al. 2003; Taylor et al. 2004; Torres et al. 1985; Wicks et al. 2000), growers have more recently begun to rely upon phosphite-based fungicides as a substitute (Johnson et al. 2004; Taylor et al. 2011).

Infected tubers used as seed material provide a source of contamination to previously pathogen-free soils (Peters and Sturz 2001). Also, *P. erythroseptica* inoculum in the soil and in infected tubers present at harvest may contaminate healthy tubers and facilitate infection through wounds made during harvest and handling operations (Salas et al. 2000; Taylor et al. 2004), ultimately resulting in additional infections during storage. Managing pink rot from harvest to storage and storage to transit is as important as controlling the pathogen during the potato growing season. In addition, growers now have the option of applying postharvest phosphite fungicides

Corresponding author: N. C. Gudmestad; E-mail: neil.gudmestad@ndsu.edu

Accepted for publication 8 October 2015.

<http://dx.doi.org/10.1094/PDIS-06-15-0696-RE>
© 2016 The American Phytopathological Society

when potato tubers are being placed into storage if they suspect they are at risk of further pink rot development. These fungicides have proven to be an extremely effective in controlling pink rot in storage when applied to tubers as a postharvest, prestorage treatment (Gachango et al. 2012a; Johnson 2008; Miller et al. 2006; Taylor et al. 2011). Growers are keenly aware that the occurrence of pink-rot-infected tubers at harvest is a strong indicator that rot will potentially develop and spread during storage. Discarding infected tubers prior to storage and managing storage conditions such as ventilation, temperature, and relative humidity are methods frequently implemented to mitigate losses in storage (Gachango et al. 2012b; Knowles and Plissey 2008). A statistical prediction of pink rot from harvest to postharvest storage based upon the level of disease observed at harvest would assist growers in deciding whether additional chemical applications are warranted, whether to expedite their marketing timeline, and whether additional storage management practices should be implemented.

Plant pathology research often encounters data (continuous response variables) bounded on the interval (0,1). Researchers use regression models to assess the response-covariate relationship and a Gaussian linear regression model is most preferred (Ferrari and Cribari-Neto 2004). These models are often based on the normal distribution of error assumption and, for percentage data, researchers often transform the response variable before a linear regression structure is fitted to the transformed data. Although transformation followed by regression is a common practice among analysts, there are newer general modeling approaches that may be more appropriate for analyzing these types of data (Schmid et al. 2013; Warton and Hui 2011). Regular linear regression models can be inadequate for situations where continuous response variables such as percentages, proportions, concentrations, and fraction values are restricted to the interval (0,1) (Bayes et al. 2012; Ferrari and Cribari-Neto 2004; Kieschnick and McCullough 2003). Linear regression models assuming a normal distribution have computational flexibility but can result in misleading conclusions for data restricted to a scale of 0 or 1 (Galvis et al. 2014). Additionally, linear regression models yield fitted values for confined percentage data (0,1) exceeding the confined domain and fail to account for asymmetries between variables (Bonat et al. 2014; Ferrari and Cribari-Neto 2004).

To overcome this, a beta regression model is proposed to model bounded (0,1) continuous response variables (Ferrari and Cribari-Neto 2004). The beta density has a flexible shape representing a variety of distributions for effective modeling of percentage outcomes confined by lower and upper limits (Bonat et al. 2014; Smithson and Verkuilen 2006). This model is widely used for analyzing percentage variables of interest for several applications and is well acknowledged (Bayes et al. 2012; Kelley et al. 2007; Schmid et al. 2013; Wallis et al. 2009). Despite its adaptability in accounting for heteroscedasticity, nonnormality, and skewness of the data, the potential of beta regression is limited to response variables supported on the interval (0,1) (Galvis et al. 2014).

The first objective of this study was to design a beta regression function to predict the development of pink rot in storage based on the incidence of the disease taken at harvest. The second objective was to provide detailed statistical functions to fit the beta regression model and discuss the rationale for this type of research.

Materials and Methods

Field plots and soil infestation. This regression study included raw data from 11 field trials conducted in North Dakota and Minnesota from 2006 to 2014. Data from multiple locations and two potato cultivars representing diverse crop growing conditions were used for regression analysis. Pink rot field trials were established near Tappen, ND (one trial), Inkster, ND (one trial), and Park Rapids, MN (nine trials). The pink rot moderately susceptible 'Ranger Russet' (Peters and Sturz 2001) and moderately resistant 'Russet Burbank' (Taylor et al. 2008) were planted in these trials. The trials were conducted to test several existing and new alternative fungicide chemistries for the control of pink rot of potato. In each trial, study plots contained a nontreated, inoculated control treatment. The accumulated data for the control

treatment were fitted to the beta regression model to predict the development of pink rot from harvest to storage.

Pink rot field inoculations were performed using a technique modified from previously described methods (Mulrooney and Gregory 2002; Taylor et al. 2011). Three isolates of *P. erythroseptica* sensitive to mefenoxam were grown on V8 media (10%) and incubated at 25°C in the dark for 4 to 6 weeks until oospores were produced. Culture plates were mechanically homogenized at the rate of 160 culture plates with 4 liters of water. At planting, the ensuing mixture (culture plate content + water) was applied in-furrow at the rate of 625 ml of inoculum slurry per 7.6 m of row. Because only the center two rows were infested in a four-row experimental unit, the inoculum rate was equivalent to approximately 25 culture plates per 7.6 m of row or one plate per 0.3 m of row.

All trials were planted in a randomized complete block design consisting of four replications per treatment and four rows per experimental unit. Depending on the research objectives and prevailing weather conditions, planting was initiated during late April to the first week of June. Seed tubers were planted in each of four rows (7.6 m long by 0.9 m wide) having a seed spacing of 0.3 m. Standard potato agronomic practices, typical of the region and crop, were implemented during the growing season. Herbicides and insecticides were applied as recommended for a commercial potato crop by the grower cooperator or our research team to manage weed and insect pests during each growing season. As needed, an overhead sprinkler irrigation system was used to irrigate all plots to maintain adequate soil moisture for plant growth during the growing season. Two days prior to harvest, vines were killed mechanically by means of a rotobearer.

Postharvest disease assessment. Tubers were harvested from two whole middle rows of each four-row plot and pink rot incidence and yield were recorded. At harvest, tubers were inspected on the harvester for visual symptoms of pink rot. Rotted tubers and tuber pieces were separated from healthy (symptomless) tubers and weighed. Pink rot incidence (percent tuber rot), where weight of healthy tubers + weight of rotten tubers = total yield, was determined as (pink-rot-infected tuber weight/total yield) × 100 = percent rot at harvest. Rotted tubers were discarded and the remaining healthy-appearing tubers harvested from only one row of the center two rows of each plot were transported to North Dakota State University for storage and postharvest evaluation. Tubers were stored in burlap bags in a walk-in cooler at 12.8°C and 85% relative humidity, conditions simulating those of commercial storage. Postharvest evaluations for pink rot incidence were performed after 32 to 78 days in storage, depending on the individual study year. During the postharvest pink rot evaluation, tuber samples were removed from the storage and the percentage of healthy and pink-rot-infected tubers was determined by weight, as previously described. Overall, 42 observations (2 missing data) from 11 study trials were used to design the beta regression prediction model.

Beta Regression model. In this study, the response variable (pink-rot-infected tubers calculated as a percentage of total tuber weight) is a continuous measurement over certain storage durations. A linear regression model is not accurate for such an analysis because the percent postharvest pink rot is analyzed on a scale having confined endpoint, (0% means no disease and 100% means maximum disease) and the response variable is measured from the 0 end of the scale to the 100 end (corresponding to 0 or 1). To overcome this problem, a beta regression model, which is a generalized linear model, was introduced (Ferrari and Cribari-Neto 2004). The probability beta density [$y \sim B(p, q)$] for dependent variable y is defined in its general form as (Ferrari and Cribari-Neto 2004):

$$f(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1}, \quad 0 < y < 1, \quad (\text{eq. 1})$$

where p and q are unknown parameters controlling the shape of the distribution, $p, q > 0$, y is a dependent variable, and $\Gamma(\cdot)$ is the gamma function.

In beta regression, it is common practice to define the two shape parameters (p, q) of density to that of the mean (μ) and precision parameter (ϕ) (Meaney and Moineddin 2014). After reparameterization to equation 1 in terms of $\mu = p/(p+q)$ and $\phi = p+q$, the probability

beta density of a random variable y with a beta distribution [$y \sim \mathbf{B}(\mu, \varphi)$] can be written as (Ferrari and Cribari-Neto 2004):

$$f(y; \mu, \varphi) = \frac{\Gamma(\varphi)}{\Gamma(\mu\varphi)\Gamma((1-\mu)\varphi)} y^{\mu\varphi-1} (1-y)^{(1-\mu)\varphi-1}, \quad 0 < y < 1, \quad (\text{eq. 2})$$

where $0 < \mu < 1$ and $\varphi > 0$. From equation 2, the mean and the variance of the random variable y were defined as $E(y) = \mu$ and $\text{Var}(y) = \mu(1-\mu)/(1+\varphi)$. For the precision parameter (φ) of a fixed estimate (mean), the higher the φ value the smaller the variance of the variable (Ferrari and Zeileis 2010).

Assuming the percentage response variables were beta distributed, a beta regression model is designed. Let y_1, y_2, \dots, y_n be a random sample from beta density $\mathbf{B}(\mu_i, \varphi)$ [$y \sim \mathbf{B}(\mu_i, \varphi)$], and then the beta regression is defined as (Cepeda-Cuervo 2015):

$$g(\mu_i) = \beta_0 + x_{i1}\beta_1 + \dots + x_{ik}\beta_k = \eta_i, \quad i = 1, \dots, n \quad (\text{eq. 3})$$

where x_{i1}, \dots, x_{ip} are the covariates, $\beta_0, \beta_1, \dots, \beta_k$ are the estimated intercept and coefficients corresponding to each covariate, η_i is the linear predictor for the i th observation, and n is the sample size. Here $g(\cdot)$ is a link function, which connects the linear predictor and the response variable. The logit link was used in our study [$g(\mu) = \log(\mu/[1-\mu])$] for beta regression.

Linear regression model. In this study, the ordinary least squares regression was applied to compare the results against the proposed beta regression model. To predict the percentage response variable from one or more explanatory variables (covariates), a linear regression model is commonly used. Usually, transformation of percentage outcome is used and a linear regression model is constructed based on the transformed response. The most commonly applied is a logit transformation and is given as:

$$y^* = \log\left(\frac{y}{1-y}\right) \quad (\text{eq. 4})$$

where y^* is the transformed value and y is the original response scaling between 0 and 1. The linear regression model of the transformed value can be written as

$$y_i = \beta_0 + x_{i1}\beta_1 + \dots + x_{ik}\beta_k \quad i = 1, \dots, n \quad (\text{eq. 5})$$

where y_i is the response variable, x_{i1}, \dots, x_{ip} are covariates, $\beta_0, \beta_1, \dots, \beta_k$ are estimated intercept and coefficients corresponding to each covariate, and n is the sample size. The linear regression model was designed separately using R software. Adjusted R^2 value and diagnostic residual plots were calculated to determine the global goodness of fit for this model.

Statistical analysis. Statistical analysis and construction of the beta regression model were performed using the *betareg* package (Ferrari and Zeileis 2010) and *faraway* package (Faraway 2014) for the R software, version 3.1.2. Three independent variables—percent pink rot at harvest (x_{harvest}), yield in metric tons per hectare (x_{yield}), and the number of days after harvest (x_{DAH}), were collected along with the percent pink rot at postharvest (y). Considering the interactions between the variables, six covariates were used to estimate parameters using the logit link function. Regression parameter estimation was performed by maximum-likelihood method. The pseudo R^2 value (squared correlation between logit-link-transformed response and linear combination of predictor variables) for the beta regression model was included in the R *betareg* package. This software package also calculates the precision parameter (φ) for beta distribution from the variance of the random variable. The values 0 and 1 were included in the model using the formula:

$$y' = \frac{y \times (n-1) + 0.5}{n} \quad (\text{eq. 6})$$

where y is the original value, $y' \in (0, 1)$ is the transformed value, and n is the sample size.

The beta regression model goodness of fit is assessed using different types of diagnostic plots, including a Pearson residual plot and

deviance residual plots. The commonly applied Pearson residual for beta regression analysis is defined as (Ferrari and Zeileis 2010):

$$r_{P,i} = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{\text{Var}}(y_i)}} \quad (\text{eq. 7})$$

Similarly, the common deviance residual is defined as:

$$r_{D,i} = a_i \sqrt{d_i} \quad (\text{eq. 8})$$

where d_i is the contribution of the i th observation to the model deviance, $a_i = 1$ if $y_i > \hat{\mu}_i$, $a_i = 0$ if $y_i = \hat{\mu}_i$, and $a_i = -1$ if $y_i < \hat{\mu}_i$.

Results

Beta Regression model. The beta regression model was designed to predict disease (dependent variable) that would develop in storage based on the level of pink rot present at harvest (Fig. 1). Results from beta regression demonstrate that the interaction between percentage of tuber rot at harvest and yield is a significant ($P < 0.0001$) predictor of the model (Table 1). The beta regression diagnostic plots for deviance and Pearson residuals were used to assess the regression assumptions (Fig. 2). The Pearson and deviance residual plots satisfied the criteria of showing no detectable pattern, which means that either Pearson or deviance residuals are independent within each other and the variance was constant (Fig. 2A and B). An estimated pseudo R^2 value of 0.56 was obtained, reflecting that the global goodness-of-fit measured for this equation was adequate. The prediction probability vs. true probability for both regression models was nearly similar (Fig. 3). The precision parameter (φ) for the beta regression model was calculated as 82.16, indicating that the variance of the observed data are low. Based on the intercept and covariate estimates from the beta regression analysis, a prediction model for analyzing pink rot over time was constructed. The equation in its general form is written as:

$$\hat{\mu} = \frac{\exp(\hat{\eta})}{1 + \exp(\hat{\eta})} \quad (\text{eq. 9})$$

where ($\hat{\mu}$) represents the predicted value based on the current data. The regression model estimates for intercept and six covariates from the present study were used to determine the linear predictor $\hat{\eta}$ from the above equation (Table 1). The $\hat{\eta}$ value changes depending on the number of study covariates used for study.

Linear regression model. Results also demonstrated that the interaction between percent pink rot at harvest and yield is a significant ($P < 0.0001$) predictor (Table 1). The diagnostic residual plot assessing the error term showed that the variance is small around *Fitted value* = -2.75 ; also, *Fitted value* > -2.75 indicates that variance of residual is not constant (Fig. 4A). The Q-Q plot indicates that the normality assumption was valid (Fig. 4B). Adjusted R^2 value of 0.49 was obtained for the linear regression model. There is no precision parameter (φ) included in normal distribution.

Discussion

This study provides a regression model for predicting pink rot from harvest to postharvest storage. Pink rot prediction during storage will assist the grower to make timely decisions between disease management and marketing. A number of regression analyses have been designed and fitted for plant pathology research data. To our knowledge, the studies reported here represent the first use of a beta regression model to predict the development of a plant disease. In this study, the beta regression model (*betareg* R software) was introduced in plant pathology for estimation of continuous response variables measured on the unit interval (0,1).

Several factors contribute to pink rot development during storage. Often, identifying a good predictor is difficult because it may consist of nonsignificant variables and ignore significant others (Copas 1983). For this analysis, focus was placed on the assumption that *P. erythrospetia* causes tuber rot prior to harvest, thereby affecting tuber yield at harvest, and that postharvest infections occur via wounds made during harvest operations (Salas et al. 2000; Taylor

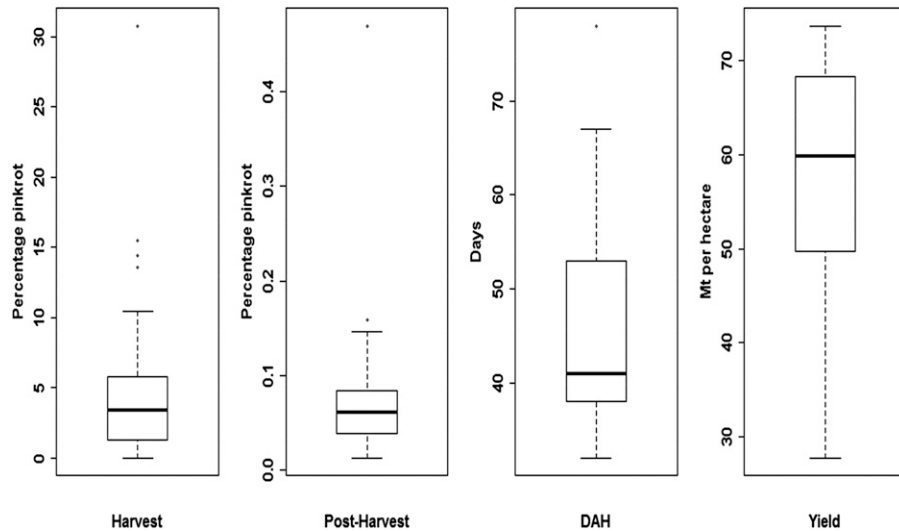


Fig. 1. Box plots for pink rot study parameters observed for tubers harvested across 11 study trials over 2006 to 2014. Dependent variable (postharvest pink rot) is predicted from independent variables harvest, days after harvest (DAH), and yield; Mt = metric ton.

Table 1. Test statistics and beta and linear regression parameter estimates^a

Coefficients	Beta Regression ^b				Linear regression ^c			
	Z test	SE	Z value	P value	t Test	SE	t Value	P value
Intercept	-5.4239	1.9629	-2.763	0.0057	-5.4419	2.1501	-2.531	0.016
Harvest	-2.3365	7.5587	-0.309	0.7572	-1.8268	9.345	-0.195	0.8461
Yield	0.0332	0.0308	1.079	0.2806	0.0321	0.0335	0.961	0.3432
DAH	0.038	0.038	1.001	0.3166	0.0446	0.0414	1.077	0.2888
Harvest × yield	0.0036	0.0008	4.737	<0.0001	0.004	0.001	3.81	<0.0001
Harvest × DAH	-0.0007	0.002	-0.349	0.7271	-0.0013	0.0025	-0.523	0.6041
Yield × DAH	-0.0006	0.0006	-0.934	0.3501	-0.0007	0.0007	-1.025	0.3123

^a SE = standard error. Harvest and days after harvest (DAH) represent percent pink rot at harvest and days after harvesting, respectively. The intercept and six coefficient estimates define the $\exp(\hat{\eta})$ for the beta regression prediction function ($\hat{\mu}$). Pseudo R^2 and adjusted R^2 represent the fitness good test for their respective regression models. No degree of freedom for Z-test estimates of beta regression, whereas $(n - 1)$ is the degree of freedom for t test estimates of linear regression.

^b Pseudo $R^2 = 0.5611$.

^c Adjusted $R^2 = 0.4934$.

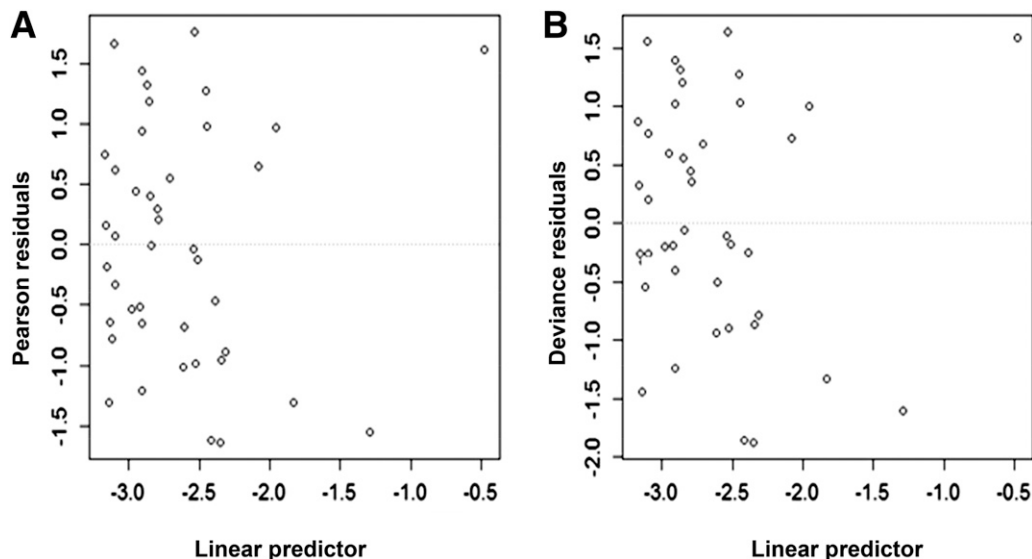


Fig. 2. Beta Regression diagnostic residual plots showing whether the errors associated with any two observations are independent in distribution. **A**, Pearson and **B**, deviance residual plots showing no pattern and considered as adequate assumption for good regression model. The regression assumption is that either Pearson or deviance residuals are independent within each other and the variance is constant.

et al. 2004). Field infections of pink rot in the United States range from 10 to 75% and, in many instances, growers avoid harvesting the highly infested areas of fields (Blodgett 1945; Boothroyd 1951; Carrol and Sasser 1974; Salas et al. 2003). However, infected tubers could be harvested from areas that are not obviously infested or infested at a very low incidence level. If not completely culled, some of these tubers will be moved into storage along with healthy tubers. Furthermore, late-season infections can occur, resulting in symptomless tubers, and tubers may become infected during harvesting operations and handling. Such infections are initiated primarily through wounds via physical contact between diseased and healthy tubers. These situations contribute to the likelihood that pink rot will manifest early during the storage period (Salas et al. 2000; Taylor et al. 2004). As rotted tubers decompose, the disease will spread and additional tubers will be infected and develop pink rot during storage.

If mefenoxam applications are part of a pink rot management regimen, the level of protection provided by this fungicide in the field

would likely be extended to tubers placed into storage as long as the local *P. erythroseptica* population remains sensitive to the fungicide and handling procedures do not cause excessive damage to the tubers as they are harvested and transported. Mefenoxam is not uniformly distributed within tuber tissue but is concentrated in biologically effective quantities at or just beneath the periderm (Barak et al. 1984; Bruin et al. 1982). Therefore skinning, cutting, and other damage to the periderm offers an unprotected, unrestricted entry point, making such tubers vulnerable to infection (Taylor et al. 2004).

Foliar applications of phosphite fungicides are highly effective in limiting pink rot tuber infections in the field (Johnson et al. 2004; Taylor et al. 2011) and as postharvest tuber treatments (Gachango et al. 2012a; Johnson 2008; Miller et al. 2006; Taylor et al. 2011). In postharvest inoculation experiments, phosphite applied to intact, nonwounded tubers prior to storage provided complete protection from infection by *P. erythroseptica* zoospores for as long as 187 days in storage, essentially rendering them immune during this period.

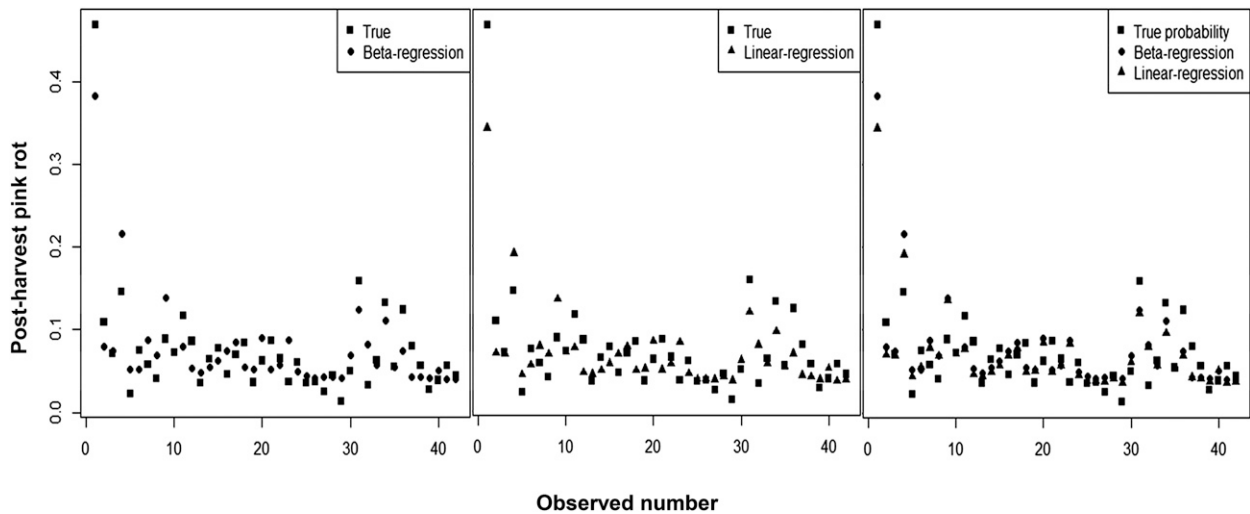


Fig. 3. Prediction probability of postharvest pink rot development compared with true probability for beta and linear regression models. For this study, predicted probability values obtained for beta and linear regression models were close to the true probability (except for a few outliers).

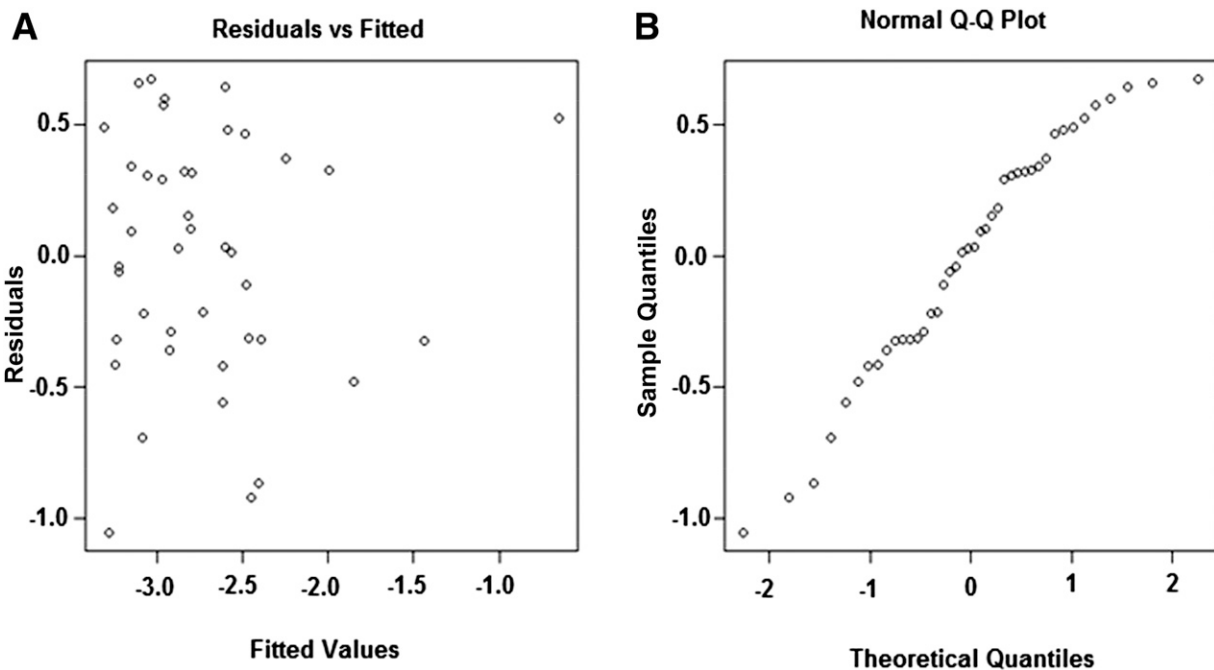


Fig. 4. Linear regression diagnostic plots for distribution of residuals were used to assess the independence assumption of error term. A, Pattern of residual plots shows that the variance is small around *Fitted value* = -2.75 and also when *Fitted value* > -2. B, Normal Q-Q plot indicates that the normality assumption is valid for linear regression.

(Taylor et al. 2004). In these experiments, phosphite protected tubers from infection by both mefenoxam-sensitive and mefenoxam-resistant isolates. Similar levels of long-term protection during storage have been reported for tubers artificially wounded and inoculated as well (Gachango et al. 2012a; Johnson 2008; Miller et al. 2006).

Prior to this study, growers did not have the means to estimate the level of rot they could expect during storage. The analysis model described here provides that tool. It offers a good estimation of rot potential if unprotected tubers are placed into storage, thereby allowing a grower to make an informed decision regarding postharvest fungicide applications. Treatment with phosphite-based fungicides would protect such tubers from infection and prevent the spread of the disease during the storage period. Any rot that ensued would most likely be directly related to quantity of rotted and symptomless tubers that were placed into storage, and the difference between the predicted rot and actual rot observed in storage would be an indicator of the efficacy of the postharvest treatment. Armed with this information, a grower could potentially adjust strategies to manage late-season infections if the proportion of this “residual rot” is high.

Yield, along with the number of days tubers are in storage, influence the final pink rot disease incidence outcome. Our study model is designed to accommodate a postharvest storage duration ranging from 32 to 78 days. Levels of “wet” storage rots exceeding 5% are generally considered unstable over the long term and must be marketed expeditiously (Secor and Gudmestad 1999). Use of the model described here could help growers to make a coordinated effort to market their crop before significant storage losses are incurred. The beta regression model is based upon observations of actual pink rot incidence in the field and, later, in storage. The model could potentially be improved in the future if highly sensitive testing for the presence of the pathogen, and not strictly observations of disease, was performed prior to placing tubers into storage. The presence of *P. erythrosetpica* associated with potato tuber tissue can be detected at extremely low levels using emerging real-time quantitative polymerase chain reaction (PCR) techniques (Atallah and Stevenson 2006). This would greatly increase the likelihood of detecting rot potential contributed by symptomless tubers and tubers carrying the pathogen as a surface contaminant. Incorporation of PCR data into future beta regression analysis could greatly enhance precision of the model. Studies such as this may further assist potato growers in making an informed decision to apply a postharvest phosphite-based fungicide as tubers are being placed into storage (Miller et al. 2006; Taylor et al. 2011).

For our study, both regression models resulted in similar predictions of postharvest pink rot development. However, the linear regression results may be biased for bounded response variables (0,1) because the heteroscedasticity is ignored with the model (Ferrari and Cribari-Neto 2004). Also, the linear regression is applied to the transformed data and, therefore, the results are interpreted based on the transformed response, which can potentially disregard the original response (Ferrari and Cribari-Neto 2004). It is not possible to directly compare the two models because the calculated pseudo R^2 of beta regression and adjusted R^2 of linear regression are statistically not similar. However, the variance pattern of residual in a diagnostic plot is not constant, indicating that the linear regression assumption may not be violated. Beta Regression was chosen for our study because its improved utility for analyzing the continuous percentage outcomes restricted to the scale (0,1) (Laliberté et al. 2012; Paolino 2001; Smithson and Verkuilen 2006). The beta distribution is flexible, accommodating left-skewed, right-skewed, and symmetric distributions of values observed on a unit interval (Kelley et al. 2007). Unlike linear regression, beta regression retains heteroscedasticity after the logit transformation because the variance continues to depend on mean and precision parameters (Saadati and Benner 2014). Also, the regression results from this model are easy to understand and coefficient parameter interpretation is very similar to logistic regression. Like normal distribution, beta distribution accounts for simultaneous modeling of the mean and precision through covariates (Grün et al. 2012). This model is simple to construct and easily accessible by researchers through standard, freely available statistical software (*betareg* for R software and GLIMMIX and FMM for SAS).

The regression model was developed using the currently available research information to predict the future response or observation ($\hat{\mu}$). If data are available for the three covariates used in our study, a grower or researcher can simply insert the appropriate values into the equation to derive an estimation of the percent development of pink rot over time ($\hat{\mu}$). However, based on the individual research goals, more predictors influencing the response variable can be added to improve or adjust the beta regression model. The linear predictor $\hat{\eta}$ from the response variable function ($\hat{\mu}$) can be used to discover predictive patterns, if present. For example, the predicted disease pattern should show an increase in pink rot development from harvest to storage. The precision parameter (φ) measures the cluster density of the observed data. The high φ value of 82.16 indicates smaller variance for clustered data near the bounds (0,1) (Schmid et al. 2013; Laliberté et al. 2012). Unlike normal distribution, where variance is high near the bounds (0,1), the φ parameter for beta distribution controls the variance and models the heteroscedasticity (Laliberté et al. 2012). For example, if $\varphi = 82.16$ and mean (μ) for the postharvest pink rot at bounds (0,1) is 0.01 or 0.99 (corresponding to 1 and 99%, respectively), the variance [$\mu(1 - \mu)/(1 + \varphi)$] is reduced to 0.0012 (very low). The residual plots (variance is constant) for our study suggest that the measured factors were important for determining the beta regression function for predicting percent pink rot during storage.

We are unaware of any other applications of beta regression in plant pathology research to estimate continuous response variables bounded on the interval (0,1). Based on the results presented here, we believe that beta regression has utility in plant pathology research as an alternative to Gaussian linear regression. This study emphasizes that the designed prediction model is adequate for analyzing pink rot over time for potato in selected cultivars grown in North Dakota and Minnesota. Because the current study used only two potato cultivars, it is possible that the current prediction function may not be accurate for other potato cultivars, particularly those that may be either highly susceptible or highly resistant to infection by *P. erythrosetpica*. Assessment of the predictive capability of the model with additional potato cultivars grown and harvested in other production areas is worth pursuing in future work. Previous studies have demonstrated that the degree of susceptibility of a potato cultivar to pink rot affects the level of chemical control that can be achieved (Taylor et al. 2008). However, this study will help plant pathologists design their own beta regression models to estimate the bounded (0,1) percentage or proportion outcomes with other disease systems. Other study covariates (if known) can be added to the existing model or a new model can be developed and tested for pink rot during storage. This model is not only used to design and develop prediction function but also is commonly applied to study how different variables influence the response variable observed on the interval (0,1). We expect that the beta regression model will be found to be useful and serve as a basis for future development in applied plant pathology research for analyzing continuous response variables such as percentages, proportions, concentrations, and fractions confined on interval (0,1).

Literature Cited

- Al-Mughrabi, K. I., Peters, R. D., Platt, H. W. (Bud), Moreau, G., Vikram, A., Poirier, R., and MacDonald, I. 2007. In-furrow applications of metalaxyl and phosphite for control of pink rot (*Phytophthora erythrosetpica*) of potato in New Brunswick, Canada. *Plant Dis.* 91:1305-1309.
- Atallah, Z. K., and Stevenson, W. R. 2006. A methodology to detect and quantify five pathogens causing potato tuber decay using real-time quantitative polymerase chain reaction. *Phytopathology* 96:1037-1045.
- Barak, E., Edgington, L. V., and Ripley, B. D. 1984. Bioactivity of the fungicide metalaxyl in potato tubers against some species of *Phytophthora*, *Fusarium*, and *Alternaria*, related to polyphenoloxidase activity. *Can. J. Plant Pathol.* 6:304-308.
- Bayes, C. L., Bazan, J. L., and Garcia, C. 2012. A new robust regression model for proportions. *Bayesian Anal.* 7:841-866.
- Blodgett, E. C. 1945. Water rot of potatoes. *Plant Dis. Rep.* 29:124-126.
- Bonat, W. H., Ribeiro, P. J., Jr., and Zeviani, W. M. 2014. Likelihood analysis for a class of beta mixed models. *J. Appl. Stat.* 42:252-266.
- Boothroyd, C. W. 1951. Pink rot of potato found in New York State. *Plant Dis. Rep.* 35:55.
- Bruin, G. C. A., Edgington, L. V., and Ripley, B. D. 1982. Bioactivity of the fungicide metalaxyl in potato tubers after foliar sprays. *Can. J. Plant Pathol.* 4:353-356.

- Cairns, H., and Muskett, A. E. 1933. Pink rot of potato. *Ann. Appl. Biol.* 20: 381-403.
- Carrol, R. B., and Sasser, M. 1974. An outbreak of potato pink rot in Delaware. *Plant Dis. Rep.* 58:443-445.
- Cepeda-Cuervo, E. 2015. Beta regression models: Joint mean and variance modeling. *J. Stat. Theory Pract.* 9:134-145.
- Copas, J. B. 1983. Regression, prediction and shrinkage. *J. R. Stat.* 45:311-354.
- Faraway, J. 2014. Functions and datasets for books by Julian Faraway. R Package version 1.0.6. Online publication. <https://cran.r-project.org/web/packages/faraway/index.html>
- Ferrari, S., and Cribari-Neto, F. 2004. Beta regression for modelling rates and proportions. *J. Appl. Stat.* 31:799-815.
- Ferrari, S., and Zeileis, A. 2010. Beta Regression in R. *J. Stat. Softw.* 34:1-24.
- Gachango, E., Kirk, W., and Schafer, R. 2012a. Effects of in-season crop-protection combined with postharvest applied fungicide on suppression of potato storage diseases caused by oomycete pathogens. *Crop Prot.* 41:42-48.
- Gachango, E., Kirk, W., Schafer, R., and Wharton, P. 2012b. Evaluation and comparison of biocontrol and conventional fungicides for control of postharvest potato tuber diseases. *Biol. Control* 63:115-120.
- Galvis, D. M., Bandyopahyay, D., and Lachos, V. H. 2014. Augmented mixed beta regression models for periodontal proportion data. *Stat. Med.* 33:3759-3771.
- Goss, R. W. 1949. Pink rot of potato caused by *Phytophthora erythroseptica* Pethyb. *Res. Bull.* 160. Online publication. University of Nebraska, College of Agriculture, Agricultural Experiment Station, Lincoln. <http://digitalcommons.unl.edu/extensionhist/1079/>
- Grün, B., Kosmidis, I., and Zeileis, A. 2012. Extended beta regression in R: Shaken, stirred, mixed, and partitioned. *J. Stat. Softw.* 48:1-25.
- Gudmestad, N. C., Taylor, R. A., and Pasche, J. S. 2007. Management of soilborne diseases of potato. *Australas. Plant Pathol.* 36:109-115.
- Johnson, D. A., Inglis, D. A., and Miller, J. S. 2004. Control of potato tuber rots caused by oomycetes with foliar applications of phosphorous acid. *Plant Dis.* 88:1153-1159.
- Johnson, S. B. 2008. Post-harvest applications of phosphorous acid materials for control of *Phytophthora infestans* and *Phytophthora erythroseptica* in potatoes. *Plant Pathol. J.* 7:50-53.
- Kelley, G., Garabed, R., Branscum, A., Perez, A., and Thurmond, M. 2007. Prediction model for sequence variation in the glycoprotein gene of infectious hematopoietic necrosis virus in California, USA. *Dis. Aquat. Organ.* 78:97-104.
- Kieschnick, R., and McCullough, B. D. 2003. Regression analysis of variables observed on (0,1): Percentages, proportions, and fractions. *Stat. Model.* 3:193-213.
- Knowles, N. R., and Plissey, E. S. 2008. Maintaining tuber health during harvest, storage, and post-storage handling. Pages 79-99 in: *Potato Health Management*, 2nd ed. D. A. Johnson, ed. American Phytopathological Society, St. Paul, MN.
- Laliberté, E., Adair, E. C., and Hobbie, S. E. 2012. Estimating litter decomposition rate in single-pool models using nonlinear beta regression. *PLoS One* 7:e45140.
- Lambert, D. H., and Salas, B. 2001. Pink rot. Pages 33-34 in: *Compendium of Potato Diseases*, 2nd ed. W. R. Stevenson, R. Loria, G. D. Franc, and D. P. Weingartner, eds. American Phytopathological Society, St. Paul, MN.
- Meaney, C., and Moineddin, R. 2014. A Monte Carlo simulation study comparing linear regression, beta regression, variable-dispersion beta regression and fractional logit regression at recovering average difference measures in a two sample design. *BMC Med. Res. Methodol.* 14:14.
- Miller, J. S., Olsen, N., Woodell, L., Porter, L. D., and Clayson, S. 2006. Post-harvest applications of zoxamide and phosphite for control of potato tuber rots caused by Oomycetes at harvest. *Am. J. Potato Res.* 83:269-278.
- Mulrooney, R. P., and Gregory, N. F. 2002. Evaluation of foliar and soil applied fungicides for the control of pink rot of potato, 2001. *Fungic. Nematicide Tests.* 57:66.
- Paolino, P. 2001. Maximum likelihood estimation of models with beta distributed dependent variables. *Polit. Anal.* 9:325-346.
- Peters, R. D., and Sturz, A. V. 2001. A rapid technique for the evaluation of potato germplasm for susceptibility to pink rot. *Plant Dis.* 85:833-837.
- Peters, R. D., Sturz, A. V., and Arsenault, W. J. 2003. Use of mefenoxam to control pink rot (*Phytophthora erythroseptica*) of potato in Prince Edward Island. *Can. J. Plant Pathol.* 25:33-40.
- Saadati, M., and Benner, A. 2014. Statistical challenges of high-dimensional methylation data. *Stat. Med.* 33:5347-5357.
- Salas, B., Secor, G. A., Taylor, R. J., and Gudmestad, N. C. 2003. Assessment of resistance of tubers of potato cultivars to *Phytophthora erythroseptica* and *Pythium ultimum*. *Plant Dis.* 87:91-97.
- Salas, B., Stack, R. W., Secor, G. A., and Gudmestad, N. C. 2000. The effect of wounding, temperature, and inoculum on the development of pink rot of potatoes caused by *Phytophthora erythroseptica*. *Plant Dis.* 84:1327-1333.
- Schmid, M., Wickler, F., Maloney, K. O., Mitchell, R., and Fenske, N. 2013. Boosted beta regression. *PLoS One* 8:e61623.
- Secor, G. A., and Gudmestad, N. C. 1999. Managing fungal diseases of potato. *Can. J. Plant Pathol.* 21:213-221.
- Smithson, M., and Verkuilen, J. 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychol. Methods* 11:54-71.
- Taylor, R. J., Pasche, J. S., and Gudmestad, N. C. 2006. Biological significance of mefenoxam resistance in *Phytophthora erythroseptica* and its implications for the management of pink rot of potato. *Plant Dis.* 90:927-934.
- Taylor, R. J., Pasche, J. S., and Gudmestad, N. C. 2008. Susceptibility of eight potato cultivars to tuber infection by *Phytophthora erythroseptica* and *Pythium ultimum* and its relationship to mefenoxam-mediated control of pink rot and leak. *Ann. Appl. Biol.* 152:189-199.
- Taylor, R. J., Pasche, J. S., and Gudmestad, N. C. 2011. Effect of application method and rate on residual efficacy of mefenoxam and phosphorous acid fungicides in the control of pink rot of potato. *Plant Dis.* 95:997-1006.
- Taylor, R. J., Salas, B., and Gudmestad, N. C. 2004. Differences in etiology affect mefenoxam efficacy and the control of pink rot and leak tuber diseases of potato. *Plant Dis.* 88:301-307.
- Taylor, R. J., Salas, B., Secor, G. A., Rivera, V., and Gudmestad, N. C. 2002. Sensitivity of North American isolates of *Phytophthora erythroseptica* and *Pythium ultimum* to mefenoxam (metalaxyl). *Plant Dis.* 86:797-802.
- Torres, H., Martin, C., and Henfling, J. 1985. Chemical control of pink rot of potato (*Phytophthora erythroseptica* Pethyb.). *Am. Potato J.* 62:355-361.
- Venkataramana, C., Taylor, R. J., Pasche, J. S., and Gudmestad, N. C. 2010. Prevalence of mefenoxam resistance among *Phytophthora erythroseptica* Pethybridge isolates in Minnesota and North Dakota. *Am. J. Potato Res.* 87: 521-530.
- Wallis, E., Mac Nally, R., and Lake, S. 2009. Do tributaries affect loads and fluxes of particulate organic matter, inorganic sediment and wood? Patterns in an upland river basin in south-eastern Australia. *Hydrobiologia* 636:307-317.
- Warton, D. I., and Hui, F. K. C. 2011. The arcsine is asinine: The analysis of proportions in ecology. *Ecology* 92:3-10.
- Wicks, T. J., Davoren, C. W., and Hall, B. H. 2000. Fungicidal control of *Phytophthora erythroseptica*: The cause of pink rot on potato. *Am. J. Potato Res.* 77:233-240.