Europ. J. Agronomy 74 (2016) 29–37

Contents lists available at ScienceDirect

European Journal of Agronomy

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

New approach to determine biological and environmental factors influencing mass of a single pea (*Pisum sativum* L.) seed in Silesia region in Poland using a CART model

Mariusz Dacko[a,*], Tadeusz Zając[b], Agnieszka Synowiec[c], Andrzej Oleksy[b], Agnieszka Klimek-Kopyra[b], Bogdan Kulig[b]

[a] Institute of Economic and Social Sciences, University of Agriculture in Krakow, Poland
[b] Institute of Plant Production, University of Agriculture in Krakow, Poland
[c] Department of Agrotechnology and Agricultural Ecology, University of Agriculture in Krakow, Mickiewicza 21, 31-120 Krakow, Poland

ARTICLE INFO

Article history:
Received 4 February 2015
Received in revised form
25 November 2015
Accepted 27 November 2015
Available online 8 December 2015

Keywords:
Yield modeling
Multivariate analysis
Regression tree
Pea variety
Pod length

ABSTRACT

There are many factors that influence the mass of a single pea seed. Because peas are an important edible legume crop of high biological value, we aimed at assessing the determinants of a single pea-seed mass based on thirteen predictors, grouped as: morphological, weather and *Rhizobium* vaccines. In a three-year field experiment, the biological and habitat characteristics of the yielding of two edible peas, i.e., semi-leafless cv. 'Tarchalska' and traditional foliage cv. 'Klif', were studied. A collection of 8446 observations was used to construct a CART regression tree. The CART model allowed us to build a regression tree with 14 terminal nodes and produced a Pearson correlation coefficient, between predicted and actual masses of seeds, of 0.71. The model satisfactorily explained the variation in the mass of a single pea-seed. It was found that the single pea-seed mass is affected mostly by 4 predictors: '*K* index', which describes the hydrothermal conditions during pea sowing-emergence (*K s-e*) and emergence-maturity (*K e-m*) periods, 'length of pod' and 'cultivar'. In the CART model, the optimal *K s-e* index and long pods were always associated with higher seed mass. The mass of a single pea-seed also depended on the 'Fruiting node'—larger seeds developed on the lowest nodes of a stem. Two factors had a marginal effect on the mass of single pea-seeds: '*Rhizobium* vaccine' and 'pod position per node'.

1. Introduction

In Europe, peas (*Pisum sativum* L.) are considered as traditional edible legumes (Borreani et al., 2007; Lopez-Bellido et al., 2005; Smil, 1997), rich in proteins of high biological value (Cousin, 1997; Davis et al., 2002). A pea is characterized by a high harvest index, in the range of 0.3–0.6 (Uzun and Açikgöz, 1998; Lecoeur and Sinclair, 2001; Jannoura et al., 2013), and a share of seed mass per pod mass in the range of 81–86% (Zając et al., 2012).

The total global area of pea cultivation is of 6.38 million hectares, and the world production of seeds is 10.98 million tons (FAO, 2013).

Pea production is highly variable as a function of different environmental conditions (Dore et al., 1998; Poggio et al., 2005). For example, the average yields of 37 recent European or Australian breeds of pea, tested for two cropping years under different climatic conditions in Italy, were significantly different between north (Lodi) –5.15 t ha^{-1} and south (Foggia) –2.52 t ha^{-1} (Annicchiarico and Iannucci, 2008).

The phenomenon of diverse pea-seeds weight is important regarding their use for food processing and direct consumption (Boros and Wawer, 2011). At cognitive and practical levels, this means that it is important to optimize plant density, which results both from the sowing plan (Dennett and Ishag, 1998), number of reproductive nodes (Jeuffroy and Devienne, 1995; Roche et al., 1998; Poggio et al., 2005) and number of seeds developed by a single plant or per unit area (canopy). It is emphasized that the total number of pea-seeds developed per unit area is an important predictor of productivity variations in specific agronomic conditions and habitats (Dore et al., 1998; Poggio et al., 2005).

Regarding the classical field experiments, it is difficult to grasp a single factor that is responsible for the formation of seed mass, as evidenced by the correlation of traits or by matrixes of corre-

* Corresponding author. Fax: +48 126624431.
E-mail addresses: m.dacko@ur.krakow.pl (M. Dacko), rrzajac@cyf-kr.edu.pl (T. Zając), a.stoklosa@ur.krakow.pl (A. Synowiec), rroleksy@cyf-kr.edu.pl (A. Oleksy), a.klimek@ar.krakow.pl (A. Klimek-Kopyra), bkulig@ar.krakow.pl (B. Kulig).

http://dx.doi.org/10.1016/j.eja.2015.11.025
1161-0301/

lation coefficients. Specifically, the issue of the mass of a single seed and the factors that determine it are ignored. For the mass of a single pea-seed, not only agro-technical factors (including the use of an appropriate *Rhizobium* vaccine) but also the habitat and weather factors may be important. The mass of a seed can also be determined by its biological origin—a particular node and pod (Atta et al., 2004). It can be assumed that both the position of a seed in a pod and its immediate neighborhood are important. It's the morphological structure of a pea pod that results in seeds developing in extreme positions in the fruit being significantly smaller than seeds located in the central part of the pod. Therefore, one of the ways to improve the yield of peas could be to search for varieties of possibly long pods, which would minimize the participation of fine seeds (Sardana et al., 2007).

As a result of a three-year field experiment (described in this paper), we received a very large amount of data related to the weight of individual pea-seeds and factors that may have a potential impact on it. Given the interconnectedness of various factors and their synergistic effects, the analysis of the results of this experiment will give better results if it is multidimensional. It was also found that the analysis should, at the same time, take into account the fact that the tested relationships do not need to be monotonic. In such cases, basic research methods, i.e., ANOVA, simple or multiple regression models, may fail. Therefore, the CART tree method was used.

The classification and regression tree (CART) method is classified, such as neural networks, as a data mining method. Models of trees, as other methods of data mining, require a considerable number—a few hundred or a few thousand—of observations to create a more comprehensive model (Breiman et al., 1998; Ferraro et al., 2009; Topal et al., 2010; Dacko and Wojewodzic, 2012). The other features of this method include its non-parametrics and non-linearity (Zheng et al., 2009; Dacko, 2010), as well as the ability to analyze non-monotonic relationships (Dacko and Wojewodzic, 2012).

The method of classification and regression trees is very versatile and can be used in any field of knowledge in which the problem is considered predictive, i.e., CART trees supported the multispectral analysis process in precision agriculture in Canada (Waheed et al., 2006) and also allowed for the extraction of factors affecting the birth weight and milk yield of cows of the 'Swedish red' cattle in Turkey (Topal et al., 2010). In agronomic research, the CART method was used to determine the factors affecting the yield of rice (Roel et al., 2007), maize (Tittonell et al., 2008), and soybean (Zheng et al., 2009). Lapen et al. (2004) used CART regression trees to examine to what extent the yield of maize in Canada is affected by the least limiting water range. Ferraro et al. (2009) used CART trees to estimate sugarcane crop yields over 5 years in Northern Argentina.

In the present study, the CART method was used to resolve which factors out of three groups most influence the weight of a single pea-seed: (1) biological position in pod and on a specific fruiting node, (2) application of *Rhizobium* vaccines and (3) weather conditions.

2. Material and methods

2.1. Brief description of the field experiment

From 2010 to 2012, a field experiment was conducted at the Experimental Station of Bayer Polska® in Modzurów (50°09′N 18°07′E, 274 m. a.s.l.), in the Silesia region, Poland. The soil of the experimental field was a degraded charnozem made of loess with a high content of available forms of phosphorus, potassium and magnesium. Pre-fertilization (kg ha^{-1}) was of: N-20, P-48 and K-72.

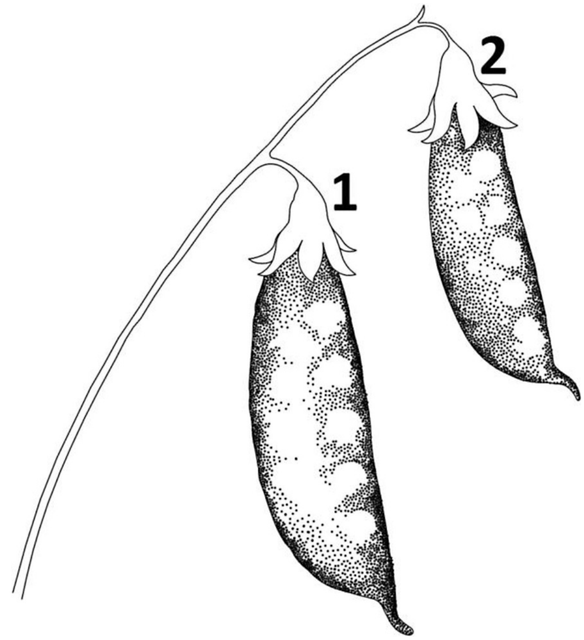

Fig. 1. Position of pod per node.

The harvested area of the pea-plot was 12 m^2, with row spacings of 15 cm. One day before sowing, pea-seeds were dressed with Funaben TTM 75 DS/WS and inoculated with *Rhizobium* vaccines: (1) a commercial product Nitragina™ (producer: BIOFOOD®) or (2) non-commercial gel-vaccine IUNG, containing *Rhizobium leguminasorum* bv. viceae—obtained from the Department of Microbiology, Institute of Soil Science and Plant Cultivation-National Research Institute in Puławy, Poland. Each year, the following amounts of seeds were sown per 1 m^2: 120 pcs. of semi-leafless pea cv. 'Tarchalska' or 100 pcs. of pea with traditional foliage cv. 'Klif'. Seeds were sown to a depth of 6 cm on April 9, 2010, April 4, 2011 and March 30, 2012.

Peas were harvested on August 6, 2010, August 3, 2011 and July 25, 2012. Shortly before harvesting, the samples of mature plants were collected for analysis of biometrics, carried out two weeks later, after drying them. From each plot, 15 shoots were analyzed regarding a series of morphological features, such as length of the fruiting part of a shoot, total length of a shoot, number of fruiting nodes per shoot, number of pods per shoot, number of seeds per pod, and mass of seeds per shoot. Moreover, the measurements of mass of a single pea-seed were made, depending on its bio-sociological position in the pod, number of seeds per pod and number of seeds per pod category–1st or 2nd (Fig. 1).

2.2. Weather conditions

During the experiment, the weather conditions were variable due to varying amounts of rainfall and air temperature during each month (Fig. 2). Season 2010 was the most humid, mainly due to heavy rainfall in the months of May and July. In subsequent years, there was less rainfall during the whole vegetation period, although in 2011, the month of July was relatively wet. In this month, shortly prior to harvesting, a whirlwind combined with rain passed through the village of Modzurów, resulting in the lodging of peas. The last year of research–2012–was relatively warm, especially the months of May and June, resulting in the accelerated vegetation of peas, which allowed for accelerated maturing and harvesting. To include the agroclimatic conditions of pea vegetation in a CART model, a hydrothermal index *K*, which combines