JOURNAL OF AGRICULTURAL INFORMATICS

AGRÁRINFORMATIKA FOLYÓIRAT

Publisher
Hungarian Association of Agricultural Informatics (HAAI)

The publishing of the Journal is supported by the European Federation for Information Technology in Agriculture, Food and the Environment (EFITA)

2018.
Vol. 9, No. 2
ISSN 2061-862X

http://journal.magisz.org

The creation of the journal has supported by the New Hungary Development

TÁMOP 4.2.3-08/1-2009-0004

"Dissemination of research result on innovative information technologies in agriculture"

The project is financed by the European Union, with the co-financing of the European Social Fund.

Nemzeti Fejlesztési Ügynökség
ÚMFT infovonal: 06 40 638 638
nfu@meh.hu • www.nfu.hu

New Hungary Development Plan

Journal of Agricultural Informatics

Scientific Journal

Name: Agricultural Informatics

Language: English

Issues: 2-4 per year

Publisher: Hungarian Association of Agricultural Informatics (HAAI), H-4032 Debrecen, Böszörményi út. 138. Hungary

ISSN 2061-862X

PREFACE

Information technology is an everyday means that is found in all walks of life today. This is also true for almost all areas of agricultural management. The aim of this Journal is to improve scientific knowledge dissemination and innovation process in the agri-food sector. The Journal of Agricultural Informatics has been established in 2009 by the HAAI within a project of the Hungarian National Development Plan Framework. The peer-reviewed journal is operating with international editorial and advisory board supported by the EFITA (European Federation for Information Technology in Agriculture Food and the Environment).

Agricultural informatics serves not only the development of the management systems of the industry but also obtaining and publicising information on production, organisation and the market for the producer.

Technologies into network based business systems built on co-operation will ensure up-to-date production and supply in food-industry. The sector-level approach and the traceability of processed agricultural products both require the application of up-to-date information technology by actors of domestic and international markets alike.

This journal serves the publication as well as familiarization the results and findings of research, development and application in the field of agricultural informatics to a wide public. It also wishes to provide a forum to the results of the doctoral (Ph.D) theses prepared in the field of agricultural informatics. Opportunities for information technology are forever increasing, they are also becoming more and more complex and their up-to-date knowledge and utilisation mean a serious competitive advantage.

These are some of the most important reasons for bringing this journal to life. The journal “Agricultural Informatics” wishes to enhance knowledge in the field of informatics, to familiarise its readers with the advantages of using the Internet and also to set up a forum for the introduction of their application and improvement.

The editorial board of the journal consists of professionals engaged in dealing with informatics in higher education, economists and staff from agricultural research institutions, who can only hope that there will be a demand for submitting contributions to this journal and at the same time there will also be interest shown toward its publications.

In this issue you can read valuable research papers comes from lecturers and co-authors of the “2017 EFITA/WCCA Congress”. The first 4 papers presented on the Congress. Thanks the authors for the interesting and useful articles and Dr. Olivier Naud (Irstea) who gave help in editing process.

Prof. Dr. Miklós Herdon

Chair of the Editorial Board

Content

Hungarian Association of Agricultural Informatics
European Federation for Information Technology in Agriculture, Food and the Environment

Journal of Agricultural Informatics. Vol. 9, No. 2
journal.magisz.org

A novel technique for fast determination of K in partitioning cluster analysis

Zeynel Cebeci[1], Cagatay Cebeci[2]

I N F O
Received 6 Mar 2018
Accepted 10 May 2018
Available on-line 18 Jun 2018
Responsible Editor: M. Herdon

Keywords:
cluster analysis,
partitioning clustering,
initialization of clustering,
number of clusters,
k selection.

A B S T R A C T

The input argument *k* refers to the number of clusters is needed to start all of the probabilistic and possibilistic partitioning algorithms. Although some progress has been made toward its solution, determining this user-specified argument is still one of the main issues in partitioning cluster analysis. Therefore, fast and even automated techniques are needed for determining *k* in partitioning clustering. In this paper, for determination of *k*, we proposed the KPEAKS, a simple and fast technique based on the descriptive statistics of peak counts of the features for clustering multidimensional datasets. The experiments on the synthetic and real datasets revealed that the mean of the largest two peak counts and the mean of third quartile and maximum peak count of the features can be successfully used for the estimates of *k*.

1. Introduction

The enormous expansion of agricultural activities and practices based on the information technologies such precision agriculture, sensory networks, RFID etc. leaded to collect the large amount of data in agriculture. Therefore data mining and big data analytics become more popular in agriculture today as well as in other areas. Clustering is one of the widely applied data mining techniques because of its usefulness in discovering the meaningful information such as the grouping structures and patterns in datasets. Clustering divides the instances in datasets into subsets called clusters by using the proximity measures (Liu *et al* 2010). According to a common taxonomy, it is possible to categorize them into three groups as hierarchical methods, partitioning methods and hybrid methods. Among them, the partitioning algorithms such as well-known K-means and Fuzzy C-means and their variants are preferred in clustering large volume of multidimensional numerical data because of their higher computational efficiencies.

Although the partitioning algorithms provide some significant advantages in clustering, they also have some disadvantages since they require a set of user-specified input arguments. However, the number and types of these arguments vary from one algorithm to another, most of the partitioning algorithms require *k*, an input argument specifying the number of partitions (or clusters) in datasets (Pakhira 2012). Using different *k* values results with different partitions, and thus, it has direct effect on the quality or validity of the final clusters. So, the choice of an appropriate value of *k* is one of the most important topics in partitioning clustering analysis (Ray & Turi 1999, Celebi *et al* 2013).

In order to determine the *k*, various subjective and objective methods have been proposed in the literature. In the subjective methods the value of *k* is determined as a priori by users. Hence, a good level of domain knowledge and experience is required with the subjective methods. On the other hand, setting it by the objective methods is mainly based on time-consuming trial and error experiments. In these experiments, a suitable clustering algorithm must be run for several times with the different values of *k*. At the end of these runs, the number of partitions which produces the best clustering result is determined by using some validity indices. Due to their computational costs, the objective methods seem impractical

[1] Zeynel Cebeci
Div. of Biometry & Genetics, Faculty of Agriculture, Çukurova University, 01330 Adana - Turkey
zcebeci@cu.edu.tr
[2] Cagatay Cebeci
Dept. of Electronic & Electrical Eng., Technology and Innovation Centre, Univ. of Strathclyde, Glasgow-UK
cagatay.cebeci@strath.ac.uk

for suggesting an optimal value of *k* especially on the real datasets that are often quite very big. Moreover, the validity indices may be sensitive to the volumes, shapes and orientations of cluster structures in the datasets.

As discussed above, deciding an optimal value of *k* is a common problem for all partitioning clustering algorithms although some progress has been made. For this reason, one of the most studied research topics on cluster analysis is on the choice of *k*. We need algorithms that will yield faster and computationally low cost solutions as datasets grow even more complex in terms of both data volume and dimensionality. Additionally, since there are differences in the information provided by the algorithms, it is not expected that the validity indices perform the same for all the clustering algorithms. Therefore, it may be necessary to use different algorithm-specific indices or robust methods that are not much influenced by cluster structures. It should be noted again that finding out a number of possible partitions and then validating them by using a validity measure is a very time consuming task. Therefore, we need the techniques giving the estimates of *k* before applying a clustering algorithm.

In this study, a novel technique, so-called "Determination of K Using Peak Counts of Features for Clustering" or shortly KPEAKS, is proposed for fast determination of *k*. The technique is based on some descriptive statistics of peak counts of the features which are found by a peaks counting algorithm. This paper is organized in different sections such that Section 2 provides the related works, Section 3 describes basic Fuzzy C-means algorithm used as a representative of partitioning clustering algorithms, Section 4 introduces the proposed technique, Section 5 discusses the performance of the proposed technique on some experimental datasets, and finally, Section 6 concludes the current study and future works.

2. Related Works

Since the partitioning algorithms produce a valid or invalid result with any value of *k*, the quality of clustering depends on the optimal choice of this input parameter. Thus, before partitioning, the number of clusters in a dataset should be determined or estimated for achieving the quality results. The value of *k* can be determined with the subjective and objective methods. In general, the subjective methods are based on heuristic approaches to understand the underlying structure of the datasets by means of various exploratory graphs (Hamerly & Elkan 2004). In this case, some degree of previous experience and domain knowledge are needed (Morissette & Chartier 2013). The subjective methods may result with poor quality clustering since the clustering algorithms may produce different results depending on the shapes and orientations of the clusters in datasets (Kodinariya & Makwana 2013). Additionally, using the subjective methods to choose *k* is exceedingly difficult and time consuming task for high dimensional data.

Objective methods mainly include the validity indices which have been primarily proposed to validate the quality of clustering results, but they can also be utilized to determine the value of *k*. These indices can be classified into three groups as the external, internal and relative indices (Kovács *et al* 2005, Rendón *et al* 2011). The external indices use some kind of external information associated with data instances. They compare the cluster labels found in a clustering analysis to the already known class labels, which can be used as the external information for deciding to an appropriate *k* value (Dudoit & Fridlyand 2002). In practice, since the external information is often not available with data, the internal validity indices are become the only applicable options. They are the validation criteria that reveal the quality of the clustering by using results obtained directly from datasets themselves (Thalamuthu *et al* 2005). Finally, the relative indices are the validity measures based on comparisons of clustering results by running one or more clustering algorithms with different input parameters on the same dataset. For instance, the best partitioning is determined by comparing the objective function values which are calculated in multiple runs of a clustering algorithm.

Cluster analysis is an unsupervised learning task in which the clustering tendencies are previously unknown. Therefore, most studies focus on the internal validity indices to validate the clustering results. These indices are generally based on the compactness, separation and their combinations. Compactness is a measure of how closely related or coherent the instances to each other. Separation, on the other hand, is a measure of how the clusters are separated from each other. There are lots of internal and

external validity indices introduced in the literature (Halkidi *et al* 2001, Rendón *et al* 2011, Charrad *et al* 2015).

There are differences in the information provided by clustering algorithms, and hence, it is not expected that all validity indices can perform in the same way in all of the clustering algorithms. For example, fuzzy and possibilistic clustering algorithms produce fuzzy membership degrees instead of crisp membership degrees, and therefore, more sophisticated internal indices may be necessary for validating their results (Wang & Zhanga 2007). Although various fuzzy indices do exist in the literature (Schwämmle and Jensen, 2010), the indices of Partition Entropy, Partition Coefficient (Bezdek 1974), Modified Partition Coefficient (Dave 1996), Xie-Beni (Xie & Beni 1991), Tang-Sun-Sun (Tang, Sun & Sun 2005), Chen-Linkens (Chen & Linkens 2004) and Pakhira-Bandyopadhyay-Maulik Fuzzy (Pakhira *et al* 2004) are often used to validate the results in fuzzy environments. These indices use membership degrees and cluster centroids obtained as a result of clustering task, and dataset itself with some indices.

In order to determine *k*, another approach tries to find the best one among all possible values with model choice via penalization by designing an appropriate penalty shape and derive an associated oracle-type inequality as proposed by Fischer (2011). The composite indices based on sensitivity and uncertainty analysis techniques, which can be used together with several cluster validity indices, have been also proposed (Marozzi 2014, Saisana *et al* 2005).

Apart from the validity indices, the information criteria such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), and some other criteria such as Minimum Description Length (MDL) and GAP statistics can also be used for determining the argument *k*. Recently, the techniques such as the Visual Assessment of Clustering Tendency (VAT) (Bezdek & Hathaway, 2002, Bezdek *et al* 2007) and an improved version of VAT (iVAT) (Havens & Bezdek 2012) have been proposed for visual determination of *k*. In addition to these, Dark Block Extraction (DBE) and Cluster Number Extraction (CCE) using the visual outputs of VAT matrices are the examples of the automated techniques for determining *k* (Pakhira 2012). Visual Assessment of Cluster Tendency Using Diagonal Tracing (VATdt) (Hu 2012) and spectral VAT (spectVAT) (Krishnamoorthi 2011) are other recently proposed algorithms in determination of *k*.

Although many validity indices are available to determine *k*, some of them are very complex to implement and some others may be computationally expensive for large datasets in many real-world applications because they require the clustering results from several runs of the algorithms. Whereas, the simpler and faster methods that can determine *k* before cluster analysis can contribute to a remarkable decrease in computational cost in partitioning cluster analysis. In Section 4 of this paper, as a new member of this kind of techniques, a novel technique enabling the fast determination of *k* is proposed.

3. Fuzzy C-means Clustering Algorithm

In the literature, the choice of *k* has mainly been worked for hard partitioning cluster algorithms with a special reference to K-means and its derivatives. In this paper, for testing the performances of the studied techniques we used the basic Fuzzy C-means Clustering (FCM) algorithm (Bezdek 1981) as the representative of partitioning clustering algorithms. As one of the most widely used soft clustering algorithms, FCM differs from hard K-means algorithm with the use of weighted squared errors instead of using squared errors only. Therefore, the proposed technique in this paper can be applied not only for FCM but also for all hard, fuzzy, possibilistic clustering algorithms and their variants in the same way. In this section, we briefly introduce the basic terminology and FCM algorithm for easy understanding the studied techniques in the paper.

Let $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n\} \in \mathbb{R}^{np}$ be a dataset to be analysed, where *n* is the number of instances, p is the number of features. For dataset $\boldsymbol{X}$, FCM tries to minimize the objective function in Eq. (1).

$$J_{FCM}(\boldsymbol{X};\ \boldsymbol{U}, \boldsymbol{V}) = \sum_{i=1}^{n} \sum_{j=1}^{k} u_{ij}^{m}\, d_{ij\boldsymbol{A}}^{2} \qquad (1)$$

The membership matrix $\boldsymbol{U}$ with $n \times k$ dimension, where *k* is the number of clusters, is a fuzzy partition of dataset $\boldsymbol{X}$ as shown in Eq. (2).

$$\boldsymbol{U} = [u_{ij}] \in M_{FCM} \qquad (2)$$

The element u_{ij} is the membership degree of ith data instance to j^{th} cluster. Thus, the jth column of matrix $\boldsymbol{U}$ contains the membership values of *n* instances to j^{th} cluster. In Eq. (3), $\boldsymbol{V}$ is a cluster prototypes matrix:

$$\boldsymbol{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_c\},\ \boldsymbol{v}_j \in \mathbb{R}^{kp} \quad (3)$$

In Eq. (1), $d^2_{ij\boldsymbol{A}}$ is the distance between i^{th} data instance and the prototype of jth cluster. It is computed using a squared inner-product distance norm in Eq. (4):

$$d^2_{ij\boldsymbol{A}} = \left\|\boldsymbol{x}_i - \boldsymbol{v}_j\right\|^2_{\boldsymbol{A}} = (\mathbf{x}_i - \mathbf{v}_j)^T \boldsymbol{A}(\boldsymbol{x}_i - \boldsymbol{v}_j) \quad (4)$$

In Eq. (4), $\boldsymbol{A}$ is a positive and symmetric norm matrix, and the inner product with norm $\boldsymbol{A}$ is a measure of distances between data points and cluster prototypes. When $\boldsymbol{A}$ is equal to $\boldsymbol{I}$, $d^2_{ik\boldsymbol{A}}$ is obtained in squared Euclidean norm. In Eq. (1), m is a fuzzifier parameter (or weighting exponent) whose value is chosen as a real number greater than one ($m \in [1, \infty)$, usually it is 2 in the literature). While m approaches to one, clustering tends to crisp like K-means but when it approaches to the infinity clustering becomes more fuzzified. The objective function J_{FCM} is minimized using the update formulas in Eq. (8) and (9) in each iteration step with the constraints in Eq. (5), (6) and (7):

$$u_{ij} \in [0,1];\ 1 \le i \le n, 1 \le j \le k \quad (5)$$

$$\sum_{j=1}^{k} u_{ij} = 1;\ 1 \le i \le n \quad (6)$$

$$0 < \sum_{i=1}^{n} u_{ij} < n;\ 1 \le j \le k \quad (7)$$

FCM stops when the iteration counts has reached to a predefined maximum iteration counts, or when the difference between the sums of membership values in $\boldsymbol{U}$ obtained two consecutive iterations is less than a predefined convergence value (ε). The steps involved in FCM are:

1. Initialize the prototype matrix $\boldsymbol{V}$ and the membership matrix $\boldsymbol{U}$.
2. Update the cluster prototypes by using Eq. (8).
$$\boldsymbol{v}_j = \frac{\sum_{i=1}^{n} u_{ij}^m \boldsymbol{x}_i}{\sum_{i=1}^{n} u_{ik}^m};\ 1 \le j \le k \quad (8)$$
3. Update the membership values by using Eq. (9).
$$u_{ij}^{(t)} = \frac{1}{\sum_{j=1}^{k} \left(d_{ij\boldsymbol{A}}/d_{lj\boldsymbol{A}}\right)^{2/(m-1)}};\ 1 \le i \le n, 1 \le j \le k \quad (9)$$
4. If $\left\|\boldsymbol{U}^{(t)} - \boldsymbol{U}^{(t-1)}\right\| < \varepsilon$ then stop else go to the step 2, where t is the iteration number.

4. Determination of K Using Peak Counts

The proposed technique, so-called "K-selection Using Peak Counts" or shortly KPEAKS, is based on some descriptive statistics of the peak counts of features by using a peaks counting algorithm. The steps involved in the technique KPEAKS are listed as follows:

1. Draw the histogram of ith feature in the dataset with the breaks which are computed by using a binning rule, i.e. Sturges and Scott or an arbitrary specified integer (Cebeci & Yildiz, 2017).
2. Run the peak finding algorithm with input arguments which are middle values and frequencies of the bins of the histogram obtained in step 1.
3. Count the peaks of ith feature in the analysed dataset, and add the obtained count into $\boldsymbol{f}$, the peak counts vector.
4. Repeat the steps 1-3 in order to count the peaks of all of the features in the dataset.
5. Calculate the descriptive statistics from the full set of peak counts in the vector $\boldsymbol{f}$.
6. Build a reduced set of peak counts by removing the peak counts smaller than a predefined threshold value of peak counts (usually 1).
7. Calculate the estimates of k on the full and reduced sets of peak counts by using the formulas in Table 1.
8. Return the list of estimates of k obtained in step 7.

As listed in Table 1, KPEAKS returns several estimates of k which are calculated by using a peak counts vector f for each feature in the analysed dataset. Some of these estimates are simply assigned from the central tendency measures without further process. For instance, the estimates $\text{KPEAKS}_{\text{AM}}$, $\text{KPEAKS}_{\text{MED}}$ and $\text{KPEAKS}_{\text{MOD}}$ are the arithmetic mean, median and mode of the peak counts, respectively. $\text{KPEAKS}_{\text{MPPC}}$ is another estimate of k which equals to the overall mean of the means of peak counts pairs. The remaining estimates of k returned by KPEAKS are calculated in different ways by using the quartiles and extreme values of the peak counts. $\text{KPEAKS}_{\text{CIQR}}$ is the centre of interquartile

range (IQR) while $KPEAKS_{CR}$ is simply the centre of range (R), or, in other words, the mean of extreme values. Finally, $KPEAKS_{MQ3M}$ is the mean of the third quartile (Q3) and maximum of peak counts, and $KPEAKS_{MTL}$ is the mean of two largest peak counts.

Table 1. KPEAKS options to determine *k*

Options	Description	Formula
AM	Arithmetic mean of peak counts	$1/p\ (\sum_{i=1}^{p} f_i)$
MPPC	Overall mean of the means of peak counts pairs	$1/\left(\frac{p^2-p}{2}\right)\left(\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}(f_i+f_j)/2\right)$
MED	Median of peak counts	$f_{(\frac{n+1}{2})}$ if n is odd else $\left(f_{(\frac{n}{2})}+f_{(\frac{n}{2}+1)}\right)/2$
MOD	Mode of peak counts	f_{mod}
CIQR	Centre of the IQR of peak counts	$1/2\left(Q3_f - Q1_f\right)$
CR	Mean of the extremes of peak counts	$1/2(f_{min}+f_{max})$
MQ3M	Mean of the Q3 and max peak count	$1/2\left(Q3_f + f_{max}\right)$
MTL	Mean of the two biggest peak counts	$1/2\left(f_{(n-1)}+f_{(n)}\right)$

*The indices between parentheses denote the order statistics of the peak counts.

Algorithm 1: **findpolypeaks**

Input:

xc, vector for the frequencies of classes of a frequency polygon

xm, vector for the middle values of classes of a frequency polygon

tc, threshold frequency value for filtering frequency polygon data, default value is 1

Output:

PM: Peaks matrix for a feature

Init:

```
xm ← xm[xc >= tc] ; xc ← xc[xc >= tc] //Filter xm and xc for the class frequencies >= tc
pfreqs ← {} //Vector for the frequencies of peaks
pvalues ← {} // Vector for the values of peaks
nc ← length of xc //Number of classes (bins)
pidx ← 1 //Index of the first peak
```

Run:

```
IF nc > 1 THEN
  IF xc[1] > xc[2] THEN
    pvalues[1]← xm[1] ; pfreqs[1]← xc[1]
    pidx ← 2
  ENDIF
  FOR i = 2 to nc-1 DO
      IF xc[i] not equal to xc[i-1] THEN
        IF xc[i] > xc[i-1] AND xc[i] >= xc[i+1] THEN
          pvalues[pidx] ← xm[i]
          pfreqs[pidx] ← xc[i]
          pidx ← pidx + 1
        ENDIF
      ENDIF
  ENDFOR
  IF xc[nc] > xc[nc-1] THEN
    pvalues[pidx]← xm[nc] ; pfreqs[pidx]← xc[nc]
  ENDIF
ELSE
  pvalues[pidx]← xm[1] ; pfreqs[pidx]← xc[1]
ENDIF
np ← length of pvalues
PMnpx2 ← 0 //Create peaks matrix
PM[,1] ← pvalues ; PM[,2] ← pfreqs
```

29: **RETURN** *PM, np*

Robustness of any estimator is important in determining *k*. It is a measure indicating the sensitivity of the estimators to the biases caused by the outliers in a dataset (Äyrämö & Kärkkäinen 2006). In this regard, $KPEAKS_{MED}$ can be considered as a robust measure of *k* because unlike $KPEAKS_{AM,}$ it is not affected by the outlying values of peak counts. $KPEAKS_{MOD}$ can also be regarded a robust metric but does not work well in multimodal cases of peak counts. When compared to $KPEAKS_{AM,}$ $KPEAKS_{MPPC}$ can provide a better estimate of *k* because it is the overall mean of the means of pairs of peak counts. As clearly seen in Figure 1, the patterns become more apparent between the features with higher peak counts. This observation shows that if estimators using higher peak counts are employed, it is possible to get more accurate estimates of *k*. Although they are not robust estimators of *k*, we could use $KPEAKS_{MQ3M}$ and $KPEAKS_{MTL}$ as useful options when the distribution of peak counts is skewed.

Finding and counting the peaks of the features in datasets are the most crucial steps in working with KPEAKS. In this paper, *findpolypeaks* (Algorithm 1), a peak finding algorithm which has been implemented in a CRAN package (Cebeci & Cebeci 2017) has been used. The input arguments of this algorithm are the frequencies (*xc*) and middle values (*xm*) of the classes of frequency polygon for the processed feature, and a threshold counts value (*tc*) for tuning the height of peaks. Here, *tc* is used for removing the little and scattered peaks formed by the outliers in analyzed datasets. The output of *findpolypeaks* algorithm are the peaks matrix (*PM*) which contains the frequency and middle values of the peaks, and peak counts (*np*) of the feature being processed.

KPEAKS can be run on the full set (FPCS) or reduced set (RPCS) of peak counts. In the first case, KPEAKS directly uses FPCS which is returned by the algorithm *findpolypeaks*. In the second case, it is applied on RPCS handled by removing the peak counts which are below a threshold level of counts from FPCS. With RPCS, it is expected that KPEAKS could produce more accurate estimates of the *k* because the features with one peak in FPCS may usually not contribute much to the formation of clustering structures.

5. Experiments on Datasets

5.1. Experiments on a Synthetic Dataset

All of the required scripts in our experiments have been implemented in R environment (R Core Team, 2018). A multidimensional synthetic dataset (dataset *5p4c*) is generated using *rnorm* function in the *stats* library of R, and it consists of five features with the descriptive statistics shown in Table 1. In the dataset consisting of 400 data instances, the first feature (*p1*) was unimodal, the second feature (*p2*) was four modal, third feature (*p3*) was three modal, fourth feature (*p4*) was four modal and fifth feature (*p5*) was bimodal.

Table 1. Descriptive statistics of the features in 5p4c dataset

Features	mean	median	std.dev.	min	max	no.peaks
p1	12.41	12.62	2.86	2.86	20.36	1
p2	69.47	69.72	22.84	29.58	109.43	4
p3	134.01	140.02	21.87	86.38	167.50	2
p4	48.67	47.83	22.57	8.70	88.58	3
p5	20.42	20.24	1.39	16.43	24.55	2

In the experiments, for computing the values of fuzzy internal indices FCM has been run for eight levels of number of clusters (k = 2,...,9). K-means++ initialization algorithm (Arthur & Vassilvitskii 2007) was used for initialization of the prototypes matrix (***V***). To avoid the possible biases due to different initializations of membership matrix (***U***), the same ***U*** matrix has been used for each level of number of clusters for the repeated runs of FCM. For this purpose, the seed of random number generator of R is set to a predefined constant number (*seed=123*). In order to validate the *k* from the results of FCM runs, some of the popular fuzzy internal indices have been used such as Partition Entropy (I_{PE}),

Modified Partition Coefficient (I_{MPC}), Xie-Beni (I_{XB}), Tang-Sun-Sun (I_{TSS}), Chen-Linkens (I_{CL}) and Pakhira-Bandyopadhyay-Maulik Fuzzy index (I_{PBMF}). In addition to the fuzzy indices listed above, the internal indices which are present in 'NbClust' package of R (Charrad *et al* 2014) have been used. Moreover, *k-selection* algorithm proposed by Pham *et al* (2005) and implemented by Rodriguez (2015) is also included because it has been argued that the algorithm is not influenced by cluster volumes. The values of all these indices have been obtained by running basic K-means algorithm with default input parameters as indicated in the package documentations. For finding the peaks of features in the analysed datasets an R implementation of Algorithm 1 have been utilized. Furthermore, an R version of the KPEAKS technique for counting the peaks and estimating the values of *k* have been coded.

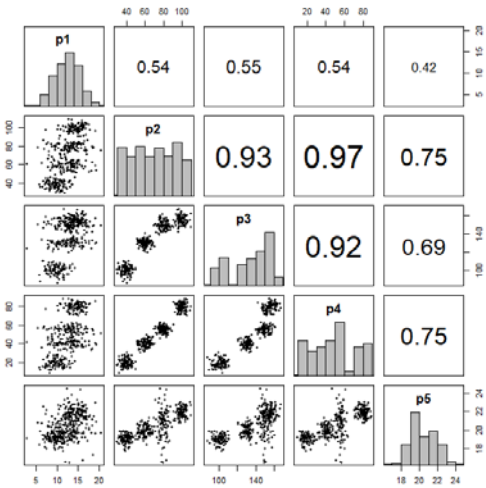


Figure 1. Histograms, scatter plots and correlations of the features in the dataset *5p4c*

In our tests, firstly the number of clusters have been estimated by using the indices in NbClust package of R (Charrad *et al* 2014). As seen in Table 2, most of the internal indices (thirteen) suggested the number of clusters as 4 for the examined synthetic dataset. Following this, five of them suggested 3 clusters, four of them suggested 2 clusters, and again two of them suggested 5 clusters. Two of the indices are evaluated as useless (i.e. Cindex proposed the number of cluster is as high as 9 while Frey proposed only 1 cluster). The k-selection algorithm suggested the number of clusters between 2 and 4 while its optimal suggestion was 2.

Table 2. Number of clusters proposed by the internal indices in NbClust

Index	***k***	**Index**	***k***	**Index**	***k***	**Index**	***k***	**Index**	***k***	**Index**	***k***
KL	4	CH	4	Hartigan	4	CCC	4	Scott	4	Marriot	4
TrCovW	3	TraceW	4	Friedman	3	Rubin	4	DB	4	Silhouette	4
Duda	3	PseudoT2	3	Beale	2	Ratkowsky	2	Ball	3	PtBiserial	2
McClain	2	Dunn	4	Hubert	4	SDindex	4	Dindex	5	SDbw	5
Frey	1	Cindex	9	kselection	2,4						

All of the studied internal fuzzy indices showed that the optimal number of clusters in the dataset *5p4c* is 4 as seen in Table 3. Since either all the fuzzy indices or majority of the indices in 'NbClust' suggested the number of clusters to be 4, this number have been used as the reference *k* value for evaluating the success of the proposed KPEAKS technique.

Table 3. Internal fuzzy index values from FCM runs on the dataset 5p4c

k	I_{XB}	I_{TSS}	I_{PBMF}	I_{CL}	I_{MPC}	I_{PE}
2	0.07089484	28.85503	1.652040e+04	0.8074789	0.6619477	0.2123218
3	0.05797912	24.20672	2.599478e+04	0.7600571	0.6953948	0.2398124
4	**0.05096023**	**22.32389**	**1.172147e+02**	**0.8545189**	**0.8119518**	**0.1512673**
5	2.17190356	936.46075	2.458972e+06	0.7093378	0.6792884	0.2863058
6	1.98144090	846.36279	2.956919e+06	0.6583026	0.6272051	0.3177349
7	1.63315099	738.42121	4.534814e+07	0.5527667	0.5452973	0.4207573
8	1.36905662	599.96624	1.151862e+07	0.5337281	0.5248174	0.4105460
9	1.28432782	570.09529	8.160711e+07	0.4440198	0.4566922	0.4869097

Peak counting function of KPEAKS have returned the peak counts vector as f ={1,4,2,3,2} by using histograms with the Sturges binning rule (Sturges 1926). The peak counts in the vector f are completely the same with the simulated numbers of the peaks which are listed in the last column of Table 1. By using descriptive statistics of the peak counts, KPEAKS proposes the estimates of k as shown in Table 4 which varies between 2 and 4. In general, these estimates are similar to those of the indices in Table 2. When the optimal value of k is concerned as 4 according the findings from the indices in Table 2 and Table 3, $KPEAKS_{MQ3M}$ and $KPEAKS_{MTL}$ are completely successful to suggest the optimal number of clusters in the dataset *5p4c*. $KPEAKS_{CR,}$ $KPEAKS_{CIQR}$ and $KPEAKS_{MPPC}$ has given the number of clusters as 3 which is the same with those from most of the indices in Table 2. On the other hand, $KPEAKS_{AM,}$ $KPEAKS_{MED}$ and $KPEAKS_{MOD}$ produce smaller estimates of k when compared to the others.

As seen in Table 4, slightly better results have been obtained on RPCS when compared to the results from FPCS. Therefore, removing of the peak counts which are equal to 1 could produce more successful results especially for the estimates with $KPEAKS_{AM,}$ $KPEAKS_{MED}$, $KPEAKS_{MOD}$ and $KPEAKS_{MPPC.}$

Table 4. Number of clusters determined with KPEAKS

Sets	AM	MED	MOD	MPPC	CIQR	CR	MQ3M	MTL
FPCS	2	2	2	2	3	3	**4**	**4**
RPCS	3	3	2	3	3	3	**4**	**4**

5.2. Experiments on the Real Datasets

For testing the performance of KPEAKS on the real data, four real datasets imported from UCI Machine Learning Repository (Lichman 2013) and one real dataset taken from a quail fattening experiment have been used. Forest type mapping training dataset (*Foresttype*) contains remote sensing data which mapped different forest types based on their spectral characteristics at visible-to-near infrared wavelengths by using the Aster satellite images (Johnson *et al* 2012). The dataset consists of 27 features and one class variable with 4 forest types. Glass dataset (*Glass*) of US Forensic Services consists of the values of 9 structural components, i.e. Na, Fe, K, etc., measured on 214 glass samples. There are 6 classes in the dataset, which can used as reference clusters or classes for test purposes. Fisher's Iris dataset (Fisher 1936) is probably one of the most widely used datasets in testing of data mining algorithms. Iris dataset (*Iris*) contains 3 classes of 50 instances each, where each class refers to an iris flower species. In this very famous data mining test dataset, one of the species classes is linearly separable while two of them are not linearly separable from each other. Quail dataset (*Quail*) contains the observations for 4 features which are carcass weight, liver weight, heart weight and gizzard weight measured at 3rd, 4th and 5th week of age of 30 Japanese quails in a fattening experiment at a research and application farm of an agricultural faculty. The dataset consists of 4 features and 1 class variable with 3 classes refers to fattening weeks. In this dataset, since the first class is linearly separable while two of them are not linearly separable from each other. Wine dataset (*Wine*) contains the results of a chemical analysis of three different wine cultivars grown in the same region in Italy. The dataset consists of 178 records with 13 features and 1 class variable with 3 classes.

Table 5 shows the *k* values determined by the studied indices and KPEAKS on the real datasets. In the second row of this table, the numbers on the left of parentheses and the numbers between parentheses stand for the suggested *k* values and the number of indices suggesting them, respectively. In the third row, the underlined numbers show the optimal *k*, and the other numbers show all of the recommended *k* values by *k-selection* algorithm. According to the results shown in Table 5, $KPEAKS_{MQ3M}$ and $KPEAKS_{MTL}$ successfully find the number of clusters for the *Foresttype* dataset. While none of the indices determines the reported number of clusters which does exist in the dataset *Glass*, $KPEAKS_{CR}$, $KPEAKS_{MQ3M}$ and $KPEAKS_{MTL}$ have given the similar results to those of the majority of indices listed in Table 2 and Table 3. Although the most of indices including the fuzzy indices propose the number of clusters as 2 for the dataset *Iris*, $KPEAKS_{CR}$, $KPEAKS_{MQ3M}$ and $KPEAKS_{MTL}$ have been more successful like those of eight of the indices in NbClust. For the dataset *Quail*, eight of the indices in NbClust propose the number of cluster as 2, and the other eight of them propose it as 3. Similarly, $KPEAKS_{CR}$, $KPEAKS_{MQ3M}$ and $KPEAKS_{MTL}$ have found the number of clusters as 3 while other options of KPEAKS have estimated it as 2. The majority of the indices in Table 2 and Table 3 suggest the number of cluster to be 2 for the dataset *Wine*. It is again 2 according to *k-selection*, however it also proposes 3 as one of the recommendations. For this dataset, the number of clusters has been determined as 3 by $KPEAKS_{AM,}$ $KPEAKS_{MED}$ and $KPEAKS_{MPPC.}$ On the other hand, $KPEAKS_{CR}$, $KPEAKS_{MQ3M}$ and $KPEAKS_{MTL}$ determine the number of cluster as 4 which has not been proposed by the other indices.

Table 5. Number of clusters determined on the real datasets

Measures	**Foresttype**	**Glass**	**Iris**	**Quail**	**Wine**
No. clusters (k)	**4**	**6**	**2,3**	**2,3**	**3**
NbClust	3(9), 2(6)	3(9), 2(5)	**2**(10), **3**(8)	**2**(8), **3**(8)	2(11), **3**(4)
k-selection	2, 3	2, 4	**2**,4,6,8,9	**2**,**3**,4,5,6	2,**3**-6, 9,12,13
I_{PE}	2	2	**2**	**2**	2
I_{MPC}	2	3	**2**	**2**	2
I_{XB}	2	3	**2**	**2**	2
I_{TSS}	2	3	**2**	**2**	2
I_{PBMF}	3	2	**2**	**2**	2
I_{CL}	2	2	**2**	**2**	2
$KPEAKS_{AM}$	3	2	**2**	**2**	**3**
$KPEAKS_{MED}$	2	2	**2**	**2**	**3**
$KPEAKS_{MOD}$	2	2	**2**	**2**	2
$KPEAKS_{MPPC}$	3	2	**2**	**2**	**3**
$KPEAKS_{CIQR}$	3	2	**2**	**2**	2
$KPEAKS_{CR}$	3	3	**3**	**3**	4
$KPEAKS_{MQ3M}$	**4**	3	**3**	**3**	4
$KPEAKS_{MTL}$	**4**	3	**3**	**3**	4

6. Conclusions

In this paper, a fast and simple technique has been proposed to estimate *k* which is an input argument of partitioning clustering algorithms. The technique so-called KPEAKS calculates the value of *k* by using various descriptive statistics of peak counts of features in datasets. Although there are several other options that the technique can offer for determining *k*, $KPEAKS_{MQ3M}$ and $KPEAKS_{MTL}$ were found to be the most successful according to majority of the findings from experiments on the synthetic and studied real datasets.

As a final conclusion, the technique KPEAKS presents not only fast choices of *k* but also provides an opportunity to work on large datasets. Instead of using computationally expensive internal indices applied to the results from many time-consuming runs of clustering algorithms, *k* is calculated very quickly with simple formulations. Hence, a significant decrease in the required computation time to work with large datasets is expected. It is assumed that the accuracy of KPEAKS can be increased by additional procedures which remove or flatten little peaks or foothills which are very close to the higher

peaks in frequency polygons. In this direction, a future study on an algorithm to remove the foothills and take only major peaks into account for increasing the efficiency of KPEAKS is within our scope.

Acknowledgement

This research was supported by the Scientific Research Projects Coordination Unit at the Cukurova University (Grant #: FBA-2017-9730).

References

Arthur, D & Vassilvitskii, S 2007, K-means++: The advantages of careful seeding. *Proc. of the 18th Annual ACM-SIAM Symp. on Discrete Algorithms*, p. 1027-1035.

Äyrämö, S & Kärkkäinen L, 2006, Introduction to partitioning based clustering methods with a robust example. *Reports of the Dept. of Math. Info. Tech. (Univ. of Jyväskylä); Series C: Software & Comp. Eng.*, C1, pp.1-36.

Bezdek, JC, 1974, Cluster validity with fuzzy sets. *J. of Cybernetics*, vol. 3, no. 3. pp. 58-73. https://doi.org/10.1080/01969727308546047

Bezdek, JC, 1981, *Pattern recognition with fuzzy objective function algorithms.* New York, Plenum. https://doi.org/10.1007/978-1-4757-0450-1

Bezdek, JC & Hathaway RJ, 2002, VAT: A tool for visual assessment of (cluster) tendency. *Proc. of IEEE Int. Joint Conf. on Neural Networks (IJCNN 02)*, May 12-17, 2002. vol. 3, pp. 2225-2230. https://doi.org/10.1109/IJCNN.2002.1007487

Bezdek, JC, Hathaway, RJ & Huband JM, 2007 Visual assessment of fuzzy clustering tendency for rectangular dissimilarity matrices. *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 5, pp. 890-903. https://doi.org/10.1109/TFUZZ.2006.889956

Cebeci, Z & Cebeci, C 2017, kpeaks: Determination of k by using peak counts of features. R package version 0.1.0. https://CRAN.R-project.org/package=kpeaks.

Cebeci Z & Yildiz, F 2017, Unsupervised discretization of continuous variables in a chicken egg quality traits dataset. *Turkish J of Agriculture-Food Science and Technology*, vol. *5*, no. 4, pp. 315-320. https://doi.org/10.24925/turjaf.v5i4.315-320.1056

Chen, MY & Linkens, DA 2004, Rule-base self-generation and simplification for data-driven fuzzy models. *Fuzzy Sets and Systems* 142: 243–265. https://doi.org/10.1016/S0165-0114(03)00160-X

Charrad, M, Ghazzali, N, Boiteau, V & Niknafs, A 2015, Package NbClust. R package version 3.0. https://CRAN.R-project.org/package=NbClust.

Charrad, M, Ghazzali, N, Boiteau, V & Niknafs, A 2014 NbClust: An R Package for determining the relevant number of clusters in a dataset. *J. Statistical Software*, vol. 61, no. 6, pp. 1-36. https://doi.org/10.18637/jss.v061.i06

Dave, RN 1996, Validating fuzzy partitions obtained through c-shells clustering. *Pattern Recognition Letters*, vol. 17, *no.* 6, pp. 613-623. https://doi.org/10.1016/0167-8655(96)00026-8

Dudoit, S & Fridlyand, J 2002, A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biology*, vol. 3, no. 7, pp. 1-21. https://doi.org/10.1186/gb-2002-3-7-research0036

Fischer, A 2011, On the number of groups in clustering. *Statistics & Probability Letters*, vol. 81, no. 12, pp. 1771–1781. https://doi.org/10.1016/j.spl.2011.07.005

Halkidi, M, Batistakis, Y & Vazirgiannis, M 2001, On clustering validation techniques. *J. of Intelligent Information Systems*, vol. 17, no. 2/3, pp. 107–145.

Hamerly, G & Elkan, C 2004, Learning the k in k-means. *Advances in Neural Information Processing Systems*, 16, Eds. S. Thrun and L.K. Saul and B. Schölkopf. pp. 281-288. MIT Press.

Havens, TC & Bezdek, JC 2012, An efficient formulation of the improved visual assessment of cluster tendency (iVAT) algorithm. *IEEE Transactions on Knowledge & Data Engineering*, vol. 24, no. 5, pp. 813-822. https://doi.org/10.1109/TKDE.2011.33

Hu, Y 2012, VATdt: Visual assessment of cluster tendency using diagonal tracing. *American J of Computational Mathematics*, vol. 2, pp. 27-41**.** https://doi.org/10.4236/ajcm.2012.21004

Kodinariya, TM & Makwana, PR 2013, Review on determining number of cluster in K-Means clustering. *Int. J of Advanced Research in Computer Science &Management Studies*, vol. 1, no.6, pp. 90-95.

Kovács, F, Legány, C & Babos, A 2005, Cluster validity measurement techniques. *6th Int. Symp. of Hungarian Researchers on Computational Intelligence*, Nov 18-19, 2005, Budapest, Hungary.

Krishnamoorthi, 2011, Automatic evaluation of cluster in unlabeled datasets. *Proc. of Int.Conf. on Information and Network Technology.* IACSIT Press, Singapore. pp. 120-124.

Liu, Y, Li, Z, Xiong, H Gao, X & Wu, J 2010, Understanding of internal clustering validation measures. *2010 IEEE Int. Conf. on Data Mining*, pp. 911-916. https://doi.org/10.1109/ICDM.2010.35

Marozzi, M 2014, Construction, dimension reduction and uncertainty analysis of an index of trust in public iInstitutions', *Quality and Quantity*, vol. 48, no. 2, pp. 939–953. https://doi.org/10.1007/s11135-012-9815-z

Morissette, L & Chartier, S 2013, The k-means clustering technique: General considerations and implementation in Mathematica. *Tutorials in Quantitative Methods for Psychology*, vol. 9, no. 1, pp. 15-24. https://doi.org/10.20982/tqmp.09.1.p015

Pakhira, MK 2012, 'Finding number of clusters before finding clusters' *Procedia Technology* vol. 4, pp. 27-37. https://doi.org/10.1016/j.protcy.2012.05.004

Pakhira, MK, Bandyopadhyay, S & Maulik, U 2004, Validity index for crisp and fuzzy clusters. *Pattern Recognition* vol. 37, no. 3, pp. 487-501. https://doi.org/10.1016/j.patcog.2003.06.005

Pham, DT, Dimov, SS & Nguyen, CD 2005, Selection of k in K-means clustering. *J of Mechanical Engineering Science*, no. 219, pp. 103-119. https://doi.org/10.1243/095440605X8298

R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Ray, S & Turi, RH 1999, Determination of number of clusters in K-Means clustering and application in colour image segmentation. *Proc. of 4th Int. Conf. on Advances in Pattern Recog. & Digital Techniques*, Calcutta, India. Narosa Publishing House, New Delhi, India, pp. 137-143.

Rendón, E, Abundez, I, Arizmendi, A & Quiroz, EM 2011, Internal versus external cluster validation indexes. *Int. J of Computers and Communications*, vol. 5, no. 1, pp. 27-34.

Rodriguez, G 2015, kselection: Selection of k in K-means clustering. R package version 0.2.0. http://CRAN.R-project.org/package=kselection.

Saisana, M, Saltelli, A & Tarantola, S, 2005, Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *J of the Royal Statistical Society: Series A (Statistics in Society)* vol. 168, no. 2, pp. 307-323. https://doi.org/10.1111/j.1467-985X.2005.00350.x

Schwämmle, V & Jensen N 2010, A simple and fast method to determine the parameters for fuzzy c–means cluster analysis. *Bioinformatics*, vol. 26, no. 22, pp. 2841–2848, doi:10.1093/bioinformatics/btq534. https://doi.org/10.1093/bioinformatics/btq534

Sturges, H 1926, The Choice of a class-interval. *J. Amer. Statist. Assoc.* vol. 21, no. 153, pp. 65-66. https://doi.org/10.1080/01621459.1926.10502161

Tang, Y Sun, F & Sun, Z 2005, Improved validation index for fuzzy clustering. *Proc. American Control Conf., 2005.* pp. 1120-1125.

Thalamuthu, A, Mukhopadhyay, I, Zheng, X & Tseng, GC 2005, Evaluation and comparison of gene clustering methods in microarray analysis. *Bioinformatics*, vol. 22, no. 19, pp. 2405-2412. https://doi.org/10.1093/bioinformatics/btl406

Wang, W & Zhanga, Y 2007, On fuzzy cluster validity indices. Fuzzy Sets and Systems, vol. 158, pp. 2095– 2117. https://doi.org/10.1016/j.fss.2007.03.004

Xie, XL & Beni, G 1991, A validity measure for fuzzy clustering. IEEE Transactions on Pattern Analysis and *Machine Intelligence*, vol. 13, no. 8, pp. 841-847. https://doi.org/10.1109/34.85677

Hungarian Association of Agricultural Informatics
European Federation for Information Technology in Agriculture, Food and the Environment

Journal of Agricultural Informatics. Vol. 9, No. 2
journal.magisz.org

Modeling the effect of olive fruit bearing percentage on Bactrocera oleae stochastic dispersion*

Romanos Kalamatianos[1], Pavlos Bouchagier[2], Markos Avlonitis [3]

I N F O
Received 16 Feb 2018
Accepted 9 Jun 2018
Available on-line 18 Jun 2018
Responsible Editor: M. Herdon

Keywords:
olive fruit fly, dispersion, simulation, olive fruit bearing

A B S T R A C T

Olive fruit fly's dispersion depends upon the olive fruit bearing percentage inside the olive grove. Previous efforts to simulate the dispersion based on the olive fruit percentage produced promising results. However the effect of olive fruit fly's percentage and fruit fly dispersion in space were not modelled correctly in some cases. In this work we improved the dispersion model to map the data obtained better. Simulated data were compared with the measured data by computing three metric indexes. For the case of 0% olive fruit the simulated data were nearly identical to the measured data, while for the case of 30% olive fruit the reproduction of the measured data was quite promising.

1. Introduction

Olive fruit fly is a parasitic insect that infests olive groves as the development of the pre-imaginal stages (egg, larva, pupa) is solely dependent on the presence of olive fruits. Therefore, if left unchecked the olive fruit fly can cause great damage in both quality and quantity of the crop (Rice 2000; Neuenschwander & Michelakis 1978). Due to its dependence on olive fruits for its reproduction it seems logical that dispersion of the adult population of the insect would be dependent from the presence of olive fruit inside the olive grove. In their experiments, Fletcher & Kapatos (1981) showed that as the percentage of olive fruit presence decreases the olive fruit fly travels longer distances. Specifically, when olive fruit flies were released in an olive grove with 30% of olive fruit production the olive fruit flies traveled an average distance of 180m in one week. However, when released in an olive grove with 0% olive fruit (no olive fruit production) the olive fruit flies traveled an average distance of over 450m in one week.

Moreover the population evolution of the olive fruit fly depends upon climate parameters (Broufas, Pappas & Koveos 2009; Johnson et al 2011; Pappas et al 2011; Tsitsipis 1977; Tsitsipis 1980; Tsitsipis & Abatzis 1980). However, other factors play also an essential role, such as the spatiotemporal evolution of the olive fruit fly's population. A model for the population evolution of the olive fruit fly incorporating the spatial dispersion of the fly was introduced by Avlonitis, Tragoudaras & Stefanidakis (2007). This was accomplished by the stochastic generalization of the logistic equation, with arbitrary initial and boundary conditions for the olive fruit fly population.

$$(1)\quad \frac{\partial p}{\partial t} = \beta p(1-p) + c\frac{\partial^2 p}{\partial x^2} + g(p)\delta p$$

where p is the population density, is the rate of increase, $c\frac{\partial^2 p}{\partial x^2}$ is the diffusion term in space and $g(p)p$ models the spatiotemporal stochasticity and $g(p)$ being the corresponding noise amplitude.

[1] Romanos Kalamatianos
Department of Informatics, Ionian University
rkalam@ionio.gr
[2] Pavlos Bouchagier
Department of Food Technology, Technological Educational Institute of Ionian Islands
pavlosbouchagier@yahoo.com
[3] Markos Avlonitis
Department of Informatics, Ionian University
avlon@ionio.gr
* This work is an extension of Kalamatianos & Avlonitis (2015)

Voulgaris et al (2013) introduced a simulation model that estimates olive fruit fly population outbreaks, with the intent to be used as a real-time alert system. Outbreaks' estimation is achieved by providing input information about olive grove location, trap data collected from the field, along with environmental data. Through their experiments, Voulgaris et al (2013) established that predicting the time when a population outbreak will occur, could help to estimate the proper time to apply population control methods.

The aforementioned model was upgraded by Kalamatianos, Avlonitis & Stravoravdis (2015) by making the dispersion model of the olive fruit fly temperature dependent, in the sense that, when temperature exceeds a predefined threshold, olive fruit flies are motionless. Furthermore, they modified the time of immature stages of the olive fruit fly, needed to complete development. The time taken for each olive fruit fly to develop from an egg to a fully grown adult became more random instead of a fixed time which was the case in the model proposed by Voulgaris et al (2013). The simulation scenarios conducted by Kalamatianos, Avlonitis & Stravoravdis (2015), revealed how the level of infestation is affected under diverse temperature sets and drifting distances in addition to the number of starting areas the initial population emerged.

Kalamatianos & Avlonitis (2015) further upgraded the aforementioned model (Kalamatianos, Avlonitis & Stravoravdis 2015) by changing the time resolution of the simulation from days to hours. Furthermore, based upon field measurements (Fletcher & Kapatos 1981), they improved the dispersion model of the olive fruit fly based on the olive fruit's percentage in the olive grove. In this work olive fruits percentage and symmetric spatial distribution of the olive fruit fly population were not modeled in detail. While for 0% olive fruit the average traveled distance was correctly reproduced, this was not the case for the 30% of olive fruit, as it was shown by Fletcher & Kapatos (1981).

In this paper, we further improve the dispersion model in order to address the aforementioned issues and to better reproduce the findings from the experiments that were conducted by Fletcher & Kapatos (1981). We compare our results with the field measurements by Fletcher & Kapatos (1981) and use three metric indexes to validate the simulated data. We also compare our model against a field experiment conducted by Rempoulakis & Nestel (2012), in which, although olive fruit fly dispersion in relation to the presence of olive fruit was not the main focus of the experiment, it is derived that for 40% olive fruit, olive fruit flies traveled in average 46.5m three days after their release.

2. Material and Methods

2.1. Simulation Model

In this subsection, we describe the simulation model, implemented with the Python programming language, employed for our experiments.

1) Input data

For a simulation the following input data must be supplied. First, the field on which the simulation will take place must be given. The information required, is the field dimensions, the location of olive groves and the olive fruit percentage of each olive grove. Following, the climate of the field must be given, namely the temperature of the field. Although, the development and activity of the olive fruit fly depends also on more environmental parameters, such as relative humidity (Broufas, Pappas & Koveos 2009; Tsitsipis & Abatzis 1980) currently only temperature is incorporated into the simulation model. The total number of temperature values must be equal to the total number of simulation steps. In addition, the total olive fruit flies present at the start of the simulation and their position inside the field are required. Finally, the total number of simulation steps must be given.

2) Population structure

We structured the population of the olive fruit fly into the following transformation stages:

1. Egg.
2. Larva. All instars are grouped into one transformation stage.

3. Pupa.
4. Immature adult. We consider all adults with undeveloped ovaries as immature.
5. Mature adult.
6. Dead.

The first three transformation stages are considered as the immobile population of the olive fruit fly, since they do not move until the adult fly emerges. When an olive fruit fly reaches its fourth transformation stage it is able to drift on the field.

3) Development

The time resolution used by the model is hourly, which means that one simulation step corresponds to one hour following the methodology by Kalamatianos & Avlonitis (2015). In order to accurately model the development of the olive fruit fly, we utilized the Degree Day model (Wilson & Barnett 1983). Various methods have been proposed to calculate Degree Day units (Wilson & Barnett 1983; Brown 2013) however, we use the method shown in Equation (2), introduced by Voulgaris et al (2013) and slightly modified by Kalamatianos & Avlonitis (2015), to calculate the accumulated degree hour units of each fly.

$$(2)\quad DH = \frac{(t_i - T_L)\left(1 - \frac{1}{1+e^{-10(t_i - T_U)}}\right)}{24}$$

where t_i is the temperature in the i-th simulation step, T_L and T_U are the lower and upper developmental thresholds, respectively, of each olive fruit fly. When an olive fruit fly reaches its Degree Day threshold it transforms to the succeeding transformation stage.

3) Oviposition

The model assumes that all olive fruit flies are females and thus can lay eggs, only when they reach the mature adult transformation stage. Each olive fruit fly can lay up to three eggs in its lifespan, up to one egg per day and exclusively during daytime hours (Johnson et al 2011; Avidov 1954). The total number of eggs was selected to account for mortality.

4) Dispersion

The speed at which an olive fruit fly drifts inside the simulated field depends on two parameters, namely, the temperature and the olive fruits' percentage. The average speed of an olive fruit fly based on the olive fruits' percentage of its current position was calculated with Equation (3) (Kalamatianos & Avlonitis (2015),

$$(3)\quad AvgSpeed(f_{x,y}) = \frac{451.8 * e^{-0.04098 * f_{x,y}}}{wh}$$

where $f_{x,y}$ is the olive fruit percentage on (x,y) coordinates of the grid and *wh* the total daytime hours in the current week.

We observed, however, that for an olive fruit percentage of 30% the dispersion model couldn't reproduce the field measurements, for the same fruit percentage, as in Fletcher & Kapatos (1981). Therefore, we modified Equation (3) from the exponential to a linear Equation (4), as shown in Figure 1, still based on the assumption that for 100% olive fruit the olive fruit fly has no preferred direction. Thus, the emerging equation is as follows:

$$(4)\quad AvgSpeed(f_{x,y}) = \frac{-4.513 * f_{x,y} + 451.8}{wh}$$

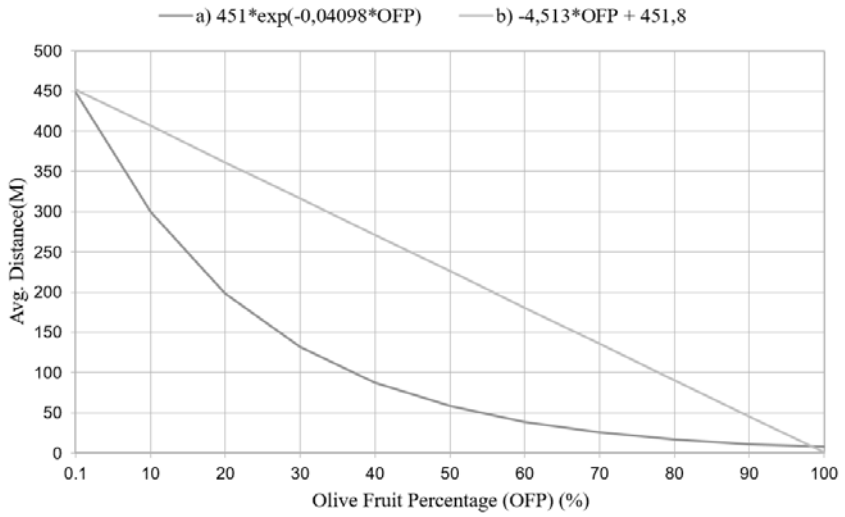


Figure 1. Curves for the mean travelled distance of olive fruit flies. Curve (a) exponential law proposed by Kalamatianos & Avlonitis (2015) curve (b) linear law proposed in this work.

The final speed is calculated by Equation (5) based on the temperature at its current position (Kalamatianos, Avlonitis & Stravoravdis, 2015).

$$(5)\ \ Speed = AvgSpeed(f_{x,y}) * \left(1 - \frac{1}{1+e^{-10*(t_i - T_M)}}\right)$$

where t_i is the temperature present in the i-th simulation step, TM is the upper movement threshold. The upper movement threshold was set to 35 ^{o}C while beyond this temperature the olive fruit flies stop drifting (Johnson et al 2011; Avidov 1954). Instead of the modified random walker model used by Kalamatianos & Avlonitis (2015), we employ alternative method for the movement of the olive fruit flies inside the grid. Specifically, each adult olive fruit fly that emerges picks a random direction to move towards. At each time step it has three choices either it continues to move towards the same direction, or it moves towards the opposite direction or it doesn't move. Each choice has a probability which changes based on the olive fruit percentage of the current position of the olive fruit fly (see Figure 2). For 100% olive fruit all choices are equiprobably.

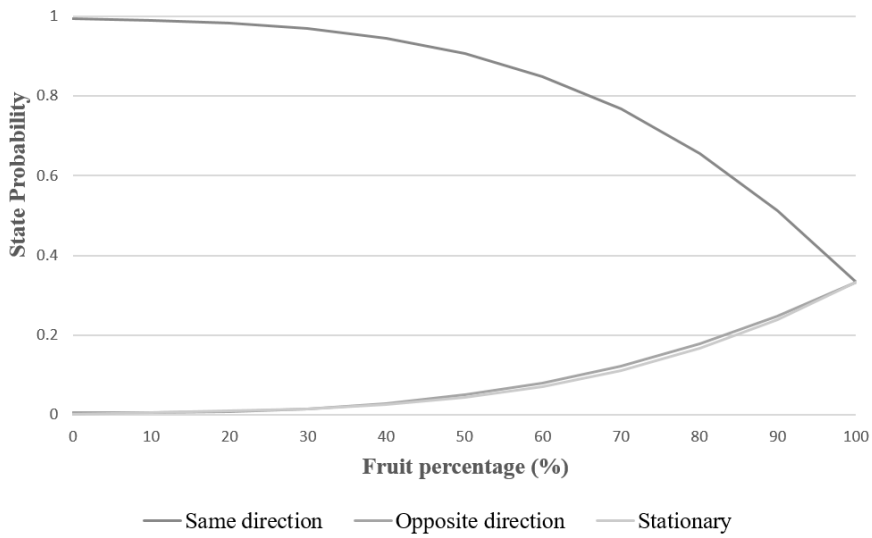


Figure 2. Probability to choose a direction state based on olive fruit percentage.

2.2. Simulation Scenarios

Herein three simulation scenarios were considered. In all scenarios, 25000 adult olive fruit flies were placed at the center of a 1500m x 1500m area olive grove. The simulation time period was set to one week which corresponds to 168 (hours) simulation time steps, for the first two scenarios. For the third

scenario the time period was set to three days or 72 simulation steps. The diurnal day/night cycle was set to 14 and 10 hours respectively. Thus, dispersion was possible in only 98 (14h daytime x 7 days) time steps for the first two simulation scenarios and 42 (14h daytime x 3 days) time steps for the last. Since, dispersion in relation to the presence of olive fruit was the main focus of the simulations, the temperature assumed was constant, namely 27°C, at all simulation steps and below the temperature movement threshold. The variable that was changed in each scenario was the olive fruit percentage, which was set to 0%, 30% and 40% respectively. Each simulation scenario was executed 100 times.

3. Results and Discussion

In this Section the results of the execution of the three simulation scenarios are presented. Table 1 - 3, all metrics were calculated with the PSPP statistical analysis software, present the results after 100 executions of all scenarios, respectively. In addition, the simulation results are compared against the measured data for validation purposes.

Table 1. 1st Simulation scenario descriptives

Distance Intervals	Mean	Std. Deviation	Variance
0-49	0.59	0.05	0.00
50-99	1.73	0.09	0.01
100-149	2.92	0.11	0.01
150-199	4.09	0.12	0.01
200-274	8.45	0.19	0.04
275-725	82.23	0.25	0.06

Table 2. 2nd Simulation scenario descriptives

Distance Intervals	Mean	Std. Deviation	Variance
0-49	17.83	0.24	0.06
50-99	24.53	0.30	0.09
100-149	21.14	0.27	0.07
150-199	15.49	0.25	0.06
200-274	12.78	0.22	0.05
275-725	8.24	0.19	0.03

Table 3. 3rd Simulation scenario descriptives

Distance Intervals	Mean	Std. Deviation	Variance
0-49	49.73	0.30	0.09
50-99	35.93	0.30	0.09
100-149	14.34	0.23	0.05
150-199	0	0	0
200-274	0	0	0
275-725	0	0	0

Figure 3 depicts the spatial distribution of the olive fruit flies for the first simulation scenario, where the olive fruits' percentage in the olive grove was 0%, which have dispersed in a circular shape. The distribution of the olive fruit flies on the outer rim is denser than the inner rim and the core of the circular shape.

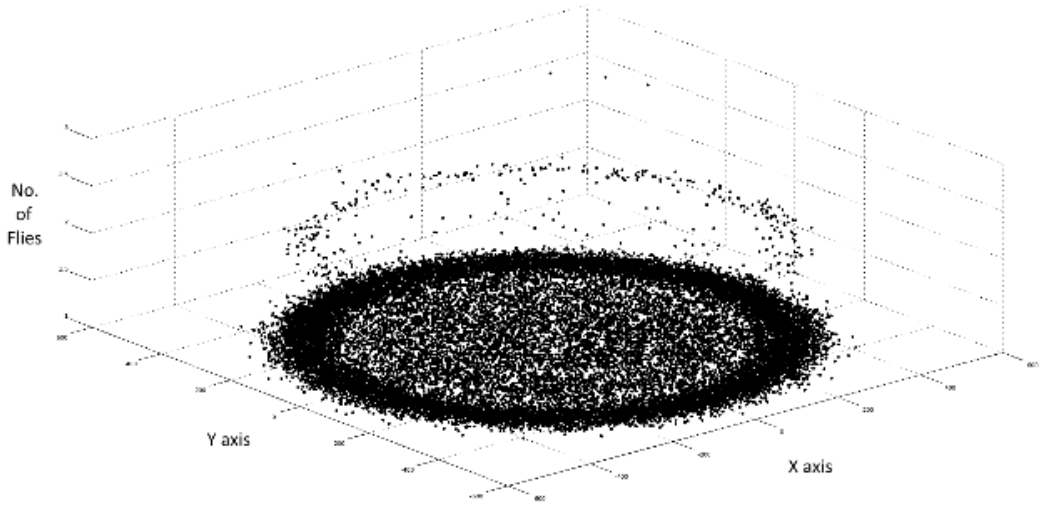


Figure 3. Spatial distribution of olive fruit flies for 0% olive fruits' presence in olive grove for a time period of one week.

Figure 4 displays the relative frequency, in percentage, of olive fruit flies at specified distance intervals after the end of the first simulation scenario. The simulated data are compared to the observed data from Fletcher & Kapatos (1981) for 0% olive fruit presence. When there is no olive fruit present in the olive grove, over 80% of the olive fruit flies have traveled a distance of at least 275m from their starting position while for distances smaller than 275m smaller percentages of olive fruit flies can be found.

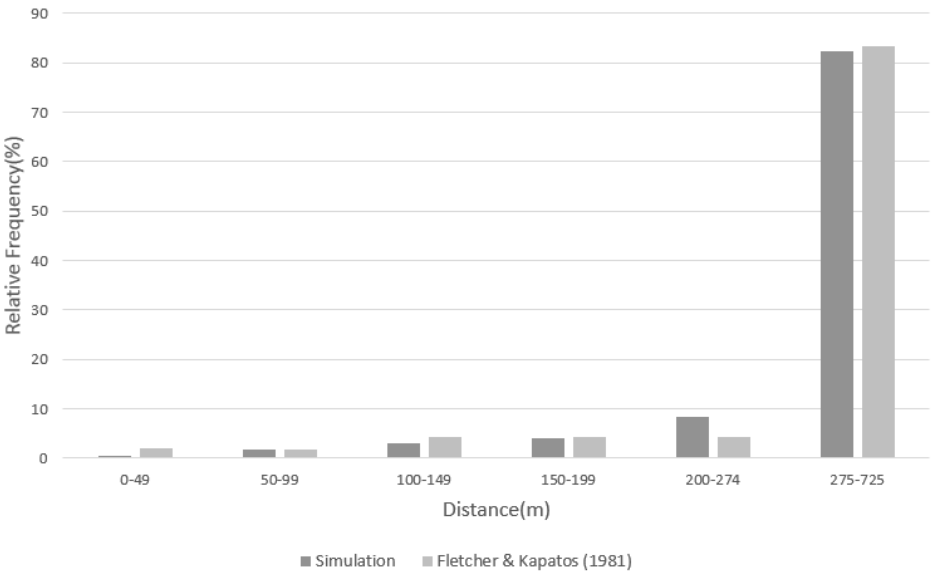


Figure 4. Relative frequency of olive fruit flies compared to observed data from Fletcher & Kapatos (1981) for 0% olive fruit presence for a time period of one week.

When compared to the observed data from Fletcher & Kapatos (1981), it seems that the simulation model reproduced the observed data quite well. Indeed, olive fruit fly presence gets more frequent as the distance from the starting position increases, while the same behavior can be seen in the observed data. A quantitative comparison of simulated and real data correlation will be given below (see Table 4).

Figure 5 depicts the spatial distribution of the olive fruit flies for the second simulation scenario, where the olive fruits' percentage in the olive grove was 30%. The spatial distribution of the olive fruit flies for the case of 30% olive fruit is the opposite of the spatial distribution for 0% olive fruits' presence. Specifically, the distribution is denser in the core and inner rim of the circular shape than the outer rim.

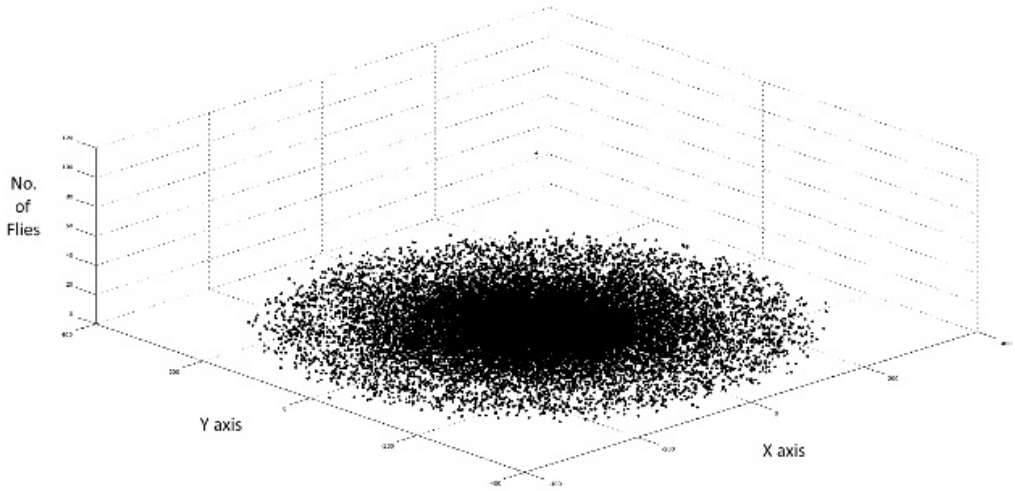


Figure 5. Spatial distribution of olive fruit flies for 30% olive fruit presence in olive grove for a time period of one week.

Figure 6 displays the relative frequency, in percentage, of olive fruit flies at specified distance intervals after the end of the second simulation scenario. The simulated data are, again, compared to the observed data from Fletcher & Kapatos (1981) for 30% olive fruit presence. The simulated data reproduce the behavior of the observed data although there are large fluctuations between the two data sets at each distance interval.

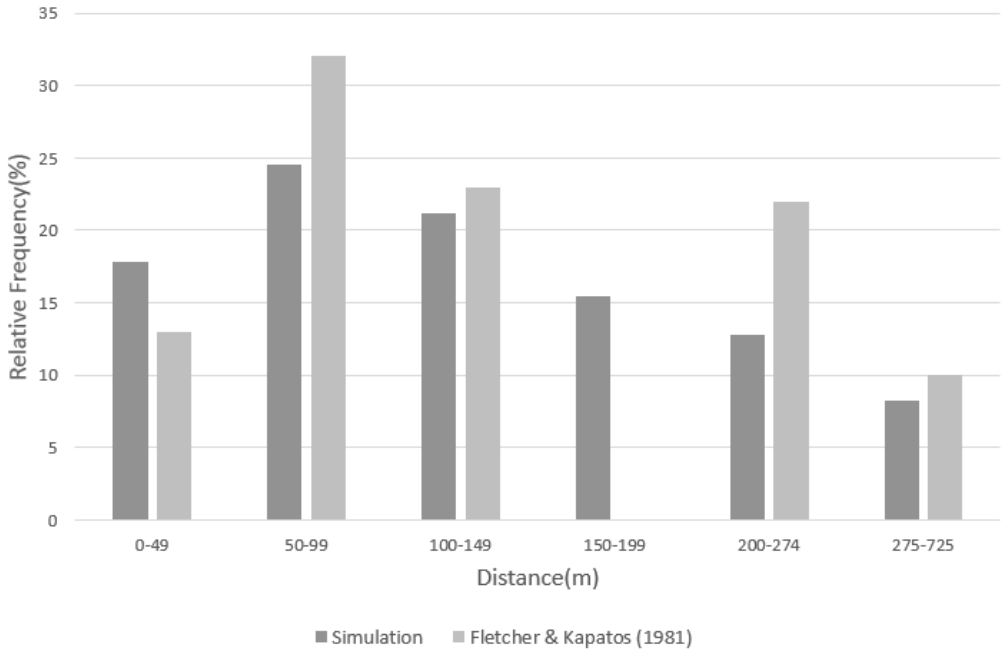


Figure 6. Relative frequency of olive fruit flies compared to observed data from Fletcher & Kapatos (1981) for 30% olive fruit presence for a time period of one week.

A comparison of the average flight distance of the olive fruit flies from the third simulation scenario compared to that measured from Rempoulakis & Nestel (2012), is shown in Figure 7. As it is shown our simulation with an average distance of 55,11m is off by approximately 5m compared to the field measurements by Rempoulakis & Nestel (2012).

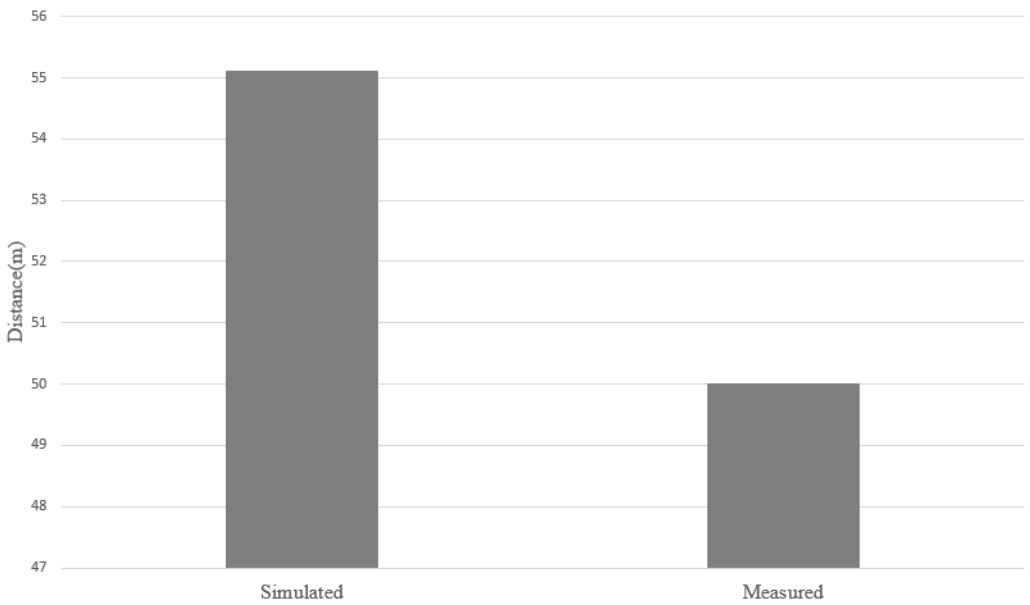


Figure 7. Average flight distance of olive fruit flies compared to observed data from Rempoulakis & Nestel (2012), for 40% olive fruit presence for a time period of three days.

To validate the dispersion model a comparison between the simulated data against the measured data from Fletcher & Kapatos (1981) was done, specifically a comparison between the relative frequencies in relation to specified distance intervals was made. The following indices, which were also used by Yang et al (2013), were used to estimate the closeness of the datasets: correlation coefficient (R), Bias (Koboyashi & Salam 2000) and root mean square error (RMSE). When the values of the two latter indices are close to zero, this indicates that the datasets are close to each other. All aforementioned indices were computed via a script in the Python programming language.

$$Bias = \frac{1}{N} \times \sum OBS_i - SIM_i$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(OBS_i - SIM_i)^2}{N}}$$

OBS_i and SIM_i are the measured and simulated value, respectively, of the ith data point in N observations.

The dispersion model validation was done for the cases of 0% and 30% olive fruit, since only for those two cases the measured data were available. Additionally, for the case of 30% olive fruit, Fletcher & Kapatos (1981) measured for the distance interval of 150-199m, no appearance of olive fruit flies, which we consider as an outlier due, for example, to trap malfunction. Therefore, we computed the aforementioned indices once for the whole dataset and once by merging the measurements for the distance interval of 150-199m with the measurements of the distance interval 200-274m. The detailed results are displayed in Table 4.

Table 4. Validation indices values

Fruit percentage	Validation indices		
	R	Bias	RMSE
0%	0.9980	1.38167	1.931
30%	0.595	6.771667	8.272802
30%*	0.775	4.438	4.999998

* merging 150 – 199m and 200-274m cases

For 0% olive fruit, the comparison gave the following results: R, Bias and RMSE were 0.9980, 1.38167 and 1.931, respectively. It is clear that the estimated relative frequency correlated well with the

measured values. However, for 30% olive fruit presence the following results were computed: 0.595, 6.771667 and 8.272802 respectively, which indicates that the estimated relative frequencies have a low correlation with the measured values. On the other hand, when we merged the distance intervals 150-199m and 200-274m, for the case of 30% olive fruit presence, the following results were computed: 0.775, 4.438 and 4.999998 respectively. By merging the aforementioned cases the estimated relative frequencies have a stronger correlation with the measured values, for this specific case.

4. Conclusion

We improved the dispersion model proposed by Kalamatianos & Avlonitis (2015) in order to better reproduce the results from the experiments conducted in Fletcher & Kapatos (1981). We validated our simulated data using the following three indexes: correlation coefficient (R), Bias (Koboyashi & Salam, 2000) and root mean square error (RMSE).

Simulation results showed, when compared to the results of Fletcher & Kapatos (1981) that for 0% olive fruit the relative frequency of olive fruit flies at specified distance intervals increases as we move further away from the origin point, which was observed as well in the field measurements, as well. Additionally, small fluctuations from the measured data were observed. When, the three metric indexes were computed they revealed a strong closeness and correlation between the simulated and measured data. For the case of 30% olive fruit presence the relative frequency of olive fruit flies at specified distance intervals have a similar tendency as the field measurements although there are large fluctuations between them. The computed values of the three metric indexes showed that there are not that strongly correlated as was the case of 0% olive fruit. Finally, when we merged, in our comparison, the values of the distance intervals 150-199m and 200-274m for the case of 30% olive fruit, since we considered that the field measurement for the 150-199m interval was a coincidence, the values of the metric indexes improved.

When compared to the results of Rempoulakis & Nestel (2012) we found that we were 9m off in regard to the average distance traveled by the olive fruit flies after three days. However, we consider this difference not too large, since in the field 5m could correspond to the distance between two olive trees.

We conclude that the proposed improvement of the olive fruit fly dispersion model, based on the comparison between the simulated and measured data, reproduced the field measurements sufficiently, especially for the case of 0% olive fruit.

Acknowledgment

The financial support of the European Union and Greece (Partnership Agreement for the Development Framework 2014-2020) under the Regional Operational Programme Ionian Islands 2014-2020, for the project “Olive Observer” is gratefully acknowledged.

References

Avidov, Z 1954, ‘Further investigations on the ecology of the olive fly (dacus oleae, gmel.) in israel’. Ktavim, vol. 4, pp. 39-50.

Avlonitis, M, Tragoudaras D & Stefanidakis, M 2007, ‘Stochastic processes and insect outbreak systems: Application to the olive fruit fly’. Proceedings of the 3rd IASME/WSEAS International Conference on energy, environment, ecosystems and sustainable development, pp. 98-103.

Broufas, G, Pappas, M & Koveos D 2009, ‘Effect of relative humidity on longevity, ovarian maturation, and egg production in the olive fruit fly (Diptera: Tephritidae)’. *Entomologia Experimentalis et Applicata*, vol. 55, pp. 161-168. https://doi.org/10.1111/j.1570-7458.1990.tb01359.x

Brown, PW 2013, ‘Heat units’. Cooperative Extension Service, University of Arizona, Tucson, Arizona. Miscell Publ.

Fletcher, BS & Kapatos, E 1981, ‘Dispersal of the olive fly, dacus oleae, during the summer period on corfu‘.*Entomologia Experimentalis et Applicata*, vol. 29, pp. 1-8. https://doi.org/10.1111/j.1570-7458.1981.tb03036.x

Johnson, M, Wang, X, Nadel, H, Opp, S, Lynn-Paterson, K, Stewart- Leslie J & Daane, K 2011, 'High temperature affects olive fruit fly populations in californias central valley'. *California Agriculture*, vol. 65, pp. 29-33. https://doi.org/10.3733/ca.v065n01p29

Kalamatianos, R & Avlonitis, M 2015, 'The role of tree distribution and olive fruit bearing in olive fruit fly infestation'. Proceedings of the 7th International Conference on Information and Communication Technologies in Agriculture, Food and Environment, Vol. 1498, CEUR Workshop Proceedings, pp. 720-729.

Kalamatianos, R, Avlonitis M & Stravoravdis, S 2015, 'Complex networks and simulation strategies: An application to olive fruit fly dispersion'. Proceedings of the 6th International Conference on Information, Intelligence, Systems and Applications (IISA), IEEE. https://doi.org/10.1109/IISA.2015.7388025

Neuenschwander, P & Michelakis, S 1978, 'Infestation of dacus oleae (gmel.) (diptera, tephritidae) at harvest time and its influence on yield and quality of olive oil in crete'. *Journal of Applied Entomology*, vol. 86, pp. 420-433. https://doi.org/10.1111/j.1439-0418.1978.tb01948.x

Pappas, ML, Broufas, GD, Koufali, N, Pieri P & Koveos, DS 2011, 'Effect of heat stress on survival and reproduction of the olive fruit fly Bactocera (Dacus) oleae'. *Journal Applied Entomology*, vol. 135, pp. 359–366. https://doi.org/10.1111/j.1439-0418.2010.01579.x

Rempoulakis, P. & Nestel, D 2012, Dispersal ability of marked, irradiated olive fruit flies [Bactrocera oleae (Rossi) (Diptera: Tephritidae)] in arid regions. Journal of applied entomology, vol. 136, no. 3, pp. 171-180. https://doi.org/10.1111/j.1439-0418.2011.01623.x

Rice, R 2000, 'Bionomics of the olive fruit fly bactrocera (dacus) oleae'. University of California Plant Protection Quarterly, vol. 10, pp. 1-5.

Tsitsipis, JA 1977, 'Effect of constant temperatures on eggs of olive fruit fly, Dacus oleae (Diptera: Tephritidae) '. *Ann. Zool. Ecol. Anim*, vol. 9, pp. 133–140.

Tsitsipis, JA 1980, 'Effect of constant temperatures on larval and pupal development of olive fruit flies reared on artificial diet'. *Environmental Entomology*, vol. 9, no.6, pp. 764–68. https://doi.org/10.1093/ee/9.6.764

Tsitsipis JA & Abatzis, C 1980, 'Relative humidity effects, at 20, on eggs of the olive fruit fly, Dacus oleae (Diptera Tephritidae), reared on artificial diet'. *Entomologia Experimentalis et Applicata*, vol. 28, pp. 92–99. https://doi.org/10.1007/BF00422028

Voulgaris, S, Stefanidakis, M, Floros A & Avlonitis, M 2013, 'Stochastic modeling and simulation of olive fruit fly outbreaks'. Procedia Technology, vol 8, 2013. https://doi.org/10.1016/j.protcy.2013.11.083

Wilson, L & Barnett WW 1983, 'Degree-days: an aid in crop and pest management'. *California Agriculture*, vol. 37, no. 1, pp. 4-7.

Yang, J, Yu H Zhao, J, Wang, Y, Wang, R & Tang L 2013, 'Mathematical Study of the Effects of Temperature and Humidity on the Morphological Development of Pleurotus Eryngii Fruit Body'. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 312-323.

Hungarian Association of Agricultural Informatics
European Federation for Information Technology in Agriculture, Food and the Environment

Journal of Agricultural Informatics. Vol. 9, No. 2
journal.magisz.org

Assessing farmers access to ICT and non-ICT sources for agricultural development in Semi-Arid Region in India

Ishwar Singh Parmar[1], Peeyush Soni[2] , Krishna R. Salin[3], John K. M. Kuwornu[4]

I N F O
Received 21 Apr 2018
Accepted 9 Jun 2018
Available on-line 18 Jun 2018
Responsible Editor: M. Herdon

Keywords:
ICT, Non-ICT, Agricultural Information, Factors affecting ICT adoption, semi-arid

A B S T R A C T

This study examines farmers' access to Information and Communication Technologies (ICT) and Non-ICT systems in the semi-arid region in the state of Rajasthan in India. The Primary data was collected from 133 farmers consisting of 68 ICT users, 62 Non-ICT users, and three moderate users. The empirical results of the multiple regression analysis revealed that education level and landholding size (in acres) influences farmers' access to ICT and non-ICT systems. However, household labor variable adversely affects access to technology. It also concludes that comparatively female farmers have lesser access to ICT. The empirical results of the multiple regression analysis revealed that education level and landholding size (in acres) positively influences farmers' access to ICT systems. However, household labor variable adversely affects access to ICT technology. It also concludes that comparatively female farmers have lesser access to ICT. Consequently, there was significant positive influence of land holding size (acres) and negative influence of education level on Non-ICT access.

The study recommends educating and sensitizing farmers about the benefits of ICT, coordination between government and private sectors for the effectual development of ICT, creating effective linkages of Krishi Vigyan Kendras (Farm Science Centers) (KVKs), Agriculture Extension as a profession for farmers and 24×7 television and radio channels dedicated to agriculture. This study can be used for productive implementation of ICT and Non-ICT sources considering the socio-economic characteristics of farmers in the similar situation.

1. Introduction

India is a growing economy with agriculture forming the backbone of the Indian economy. Despite the concentration of industrialization, agriculture remains in a place of pride (Kumar and Sankarakumar, 2012). Agriculture in the western arid region of Rajasthan is mainly rain fed. Drought, insufficient rainfall, and dry soil are the characteristics of this region. Lack of information coupled with factors like environment affect yield, quality, and price as a result farmers suffer.

ICT make a significant contribution to economic growth of agrarian sector by empowering farmers with modern technologies and creating new employment opportunities (Malhan et al., 2007). In India, KVKs and Agricultural Extension Division provide agricultural training to farmers. For sustainable growth, ICT system should be implemented in the agriculture and allied sectors as it will not only enhance traditional farming but farmers will also benefit by adopting new technologies. It has been

[1] Ishwar Singh Parmar
Asian Institute of Technology: Bangkok, Thailand
parmaris@hotmail.co.uk
[2] Peeyush Soni
Asian Institute of Technolgy, Thailand
soni@ait.asia
[3] Krishna R. Salin
Asian Institute of Technolgy, Thailand
salinkr@ait.asia
[4] John K. M. Kuwornu
Asian Institute of Technolgy, Thailand
jkuwornu@ait.asia

comprehended that the implementation of ICTs in agriculture extension would help in sensitizing farmers about the relevant information.

Agrarian community prefers traditional communication channels such as radio, television, and newspapers instead of adopting ICT (Rechandson, 2006; IICD, 2006; Hayrol et al., 2009). One of the major impediments hampering adaptation of ICT is language (Telg et al., 2005) as most of the farmers in the western Rajasthan are illiterate. Majority of the ICT application runs in English; therefore, it is challenging for the central and the state governments to empower farmers through ICT and equip them to reap the benefit of services. For the successful implementation of ICT services, government intervention is necessary to synchronize and govern the efforts of the private sector that make infrastructure and development investments. In short, Indian agriculture can improve dramatically by endorsing ICT. All farmers, including small and marginal, can be benefited.

The objectives of this study are:

1. Find socio-economic characteristics of the respondents
2. Analyze the influence of education of each respondent and employment type
3. How education influences ICT adoption
4. Find effective ICT and Non-ICT sources to assess agricultural information
5. Find factors influencing adoption of ICT sources and percentage of Non-ICT sources accessed

To meet their farming needs, farmers source agricultural knowledge from varied sources (Rees et al., 2000; Stefano et al., 2005; Karamagi Akiiki, 2006). Some are ICT and some are Non-ICT sources.

ICT Sources

Heeks (1999) elucidated ICT as “recording, tackling, storing and disseminating” information through computers and communicating systems. ICTs mean implementation of technological advancement and modernization. The ICT sources are those which accommodate information services and disseminate knowledge through steady modes such as articles, news reports, and e-mail communications.

ICT sources encourage people to communicate competently, overcome obstacles of time and space, empower individuals through knowledge and information, learn revenue-generating skills, strengthen government capabilities and encourage active participation of masses in decision-making (Asian Development Bank, 2004). ICT can be a key enabler in harmonizing the supply chain by ensuring real-time knowledge dissemination across the chain, augmenting performance and curbing unpredictability of the supply chain (Pereira, 2009).

Traditional communication used to spread agricultural knowledge include radio, television, video, fax and telephone. Television (Best et al., 2005; Stefano et al., 2005; Kwake, 2007) and Radio (Rivera et al., 2005; Byamugisha et al., 2008) are envisioned as useful communication channels. Radio is seen as one of the most effective media to share knowledge and empower marginal and illiterate farmers (Chapman et al., 2003; Harris, 2004; Best et al., 2005).

Television is also considered a vital source for circulating agricultural information (Leach, 2001; Chapman et al., 2003). The TV ads enables the audience to notice and understand the message, thus, making the medium more effective (Leach, 2001). The study by Best et al. (2005) in Bangladesh considerably judged TV and highlighted hurdles such as insufficient TV sets and inadequate supply of electricity in rural areas. Even though in the past rural societies challenged television’s capacity to disseminate information (Leach, 2001).

The video is also an effective medium of sharing agricultural knowledge in rural areas. This medium is flourishing as it overcomes illiteracy barrier and exhibits compassion while sharing modern agriculture techniques with the viewers (Colle and Roman, 2003).

Studies have manifested that the print media (books, advertisements, pamphlets, handbooks, newspaper, and leaflets) is beneficial for transferring agricultural knowledge to farmers (Stefano,

2004; Stefano et al., 2005; Klein, 2009). The print media is considered more sensible as compared to broadcast media. However, most of the print materials are not easily accessible to the farmers because these materials are primarily in English language and several authors have rued about language restriction (Leach, 2001).

ICT should upgrade and use new means such as mobile and Internet-based services to share information. It has been observed that modern ICTs have a better prospect in agricultural information (Mehra et al., 2004; Rivera et al., 2005; Gray, 2010). The e-Choupal was initiated to enhance market accessibility and knowledge (Qiang et al., 2009).

The Agriculture Extension organizes meetings, group discussions, lectures, workshops, conferences and regional training sessions utilizing ICT resources such as computers, slides and snapshots (Isife and Ofuoku, 2008). It has been observed that computers used by the extension specialists have played a vital role in the growth of the extension (Martin et al., 2001). The implementation of digital podium and ICT applications in Agricultural Extension have entailed various benefits such as boosting learning operation, improving retention time and cementing the beginners' treading (Meera et al., 2004; Park et al., 2007).

ICTs can be a counterpart of the conventional extension plan for transferring "Knowledge Resource" to the millions of the farmers (Koehnen, 2011).

Non- ICT Sources

The non-ICT sources impart information through training, extension specialists, government exhibitions, KVKs, village exhibitions, modern farmers, study visits, output dealers/commission agents, private advisors, relatives, friends, and others. In most of the developing nations, farmers do not have faith in extension specialists because of the representative's lack of knowledge and skills in modern farming techniques (Dutta, 2009).

The non-ICT sources are not equipped to provide knowledge services and communicate information verbally. Non-ICT sources comprise interactions with friends, other farmers, relatives, input dealers and output traders (Steve et al., 1999; Just et al., 2002; Mittal, 2013). Most of the farmers advocate non-ICT interpersonal sources (Rees et al., 2000; Solano et al., 2003; Stefano et al., 2005). Farmers encourage verbal communication due to conventions and low literacy rate (Meyer, 2000; Leach, 2001; Stefano et al., 2005). Apart from that, non-ICT sources of agricultural information are rural elders and community leaders (Alewis, 2008), cooperatives (Neubert et al., 2007; Chambo, 2009), local and bigger markets (Pokhrel and Thapa, 2007; Byamugisha et al., 2008), corporations (Chisenga et al., 2007), private federations and agencies (Rees et al., 2000; Klerks, 2009).

Factors Affecting Adoption of ICT-based Agricultural Information

According to Michiels and Van Crowder (2001), in comparison to conventional communication, acceptance of ICT have prospects of two-way and horizontal discourse and gaining latest communication routes for rural areas, intermediaries and development enterprises. With the acceptance of ICT sources, Agricultural Extension is envisioned to become knowledge intensifier, broader and demand-driven and thus more fruitful in meeting farmers' requirement for information (Gelb et al., 2008).

Four prerequisites for the sustainability of ICT hubs and acceptance of ICT by the locals are debated. They are the viability of funds, recognition, the staff's potential and distribution of services (Harris, 2004).

It was found that the consequence of aging is not clear for the acceptance of ICT (Putler and Zilberman, 1988; Gibbon and Warren, 1992; Warren et al., 1996), still several research technologists found a contradictory impact of age on acceptance of ICT (Batte et al., 1990; Warren et al., 2000). The adoption of ICT is interrelated to education and farm dimension of farmers (e.g. Putler and Zilberman, 1988; Batte et al., 1990; Bonny, 1992; Gibbon and Warren, 1992; Warren et al., 2000).

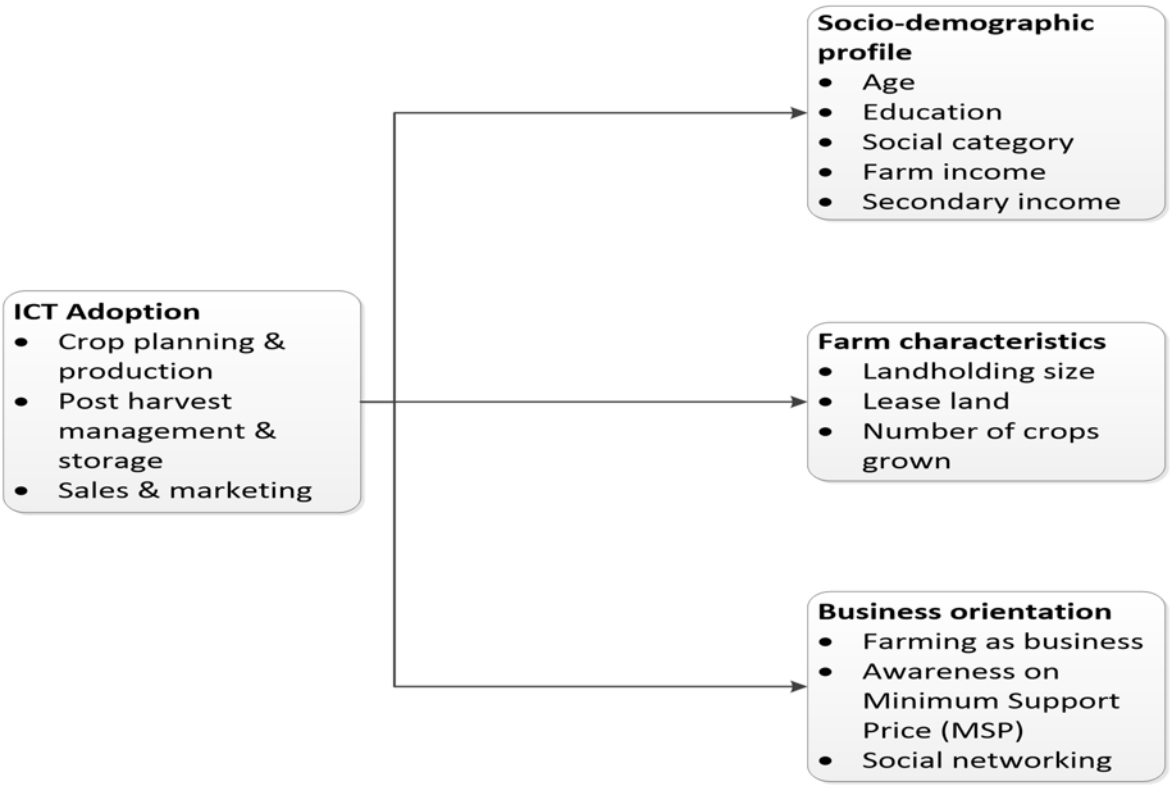


Figure 1 Conceptual framework- factors affecting adoption of ICT based agricultural information

Source: Redrawn after Jabir Ali (2012)

2. Materials and Methods

2.1. Area of Study

The study was conducted in the semi-arid region of the state of the Rajasthan in India during May 2016 to July 2016 to evaluate the accessibility of ICT and Non-ICT sources for agricultural information. The study used the survey approach consisting of various data collection strategies, including field task, documents, and findings.

2.2. Sampling and Data

The sample households for data collection were chosen using multistage sampling technique. In the first stage, a goal-directed selection was adopted to choose three districts - Jalore, Pali, and Sirohi– in the south-west region of Rajasthan. A simple random sampling was chosen to collect samples of farm households from each district. In this research, out of 133 randomly selected farm households from three districts, 3 were moderate users, 68 were ICT users, and 62 were Non-ICT.

2.3. Methods of Analysis

The research is predominantly based on primary data collected from farm households. Relevant secondary data from trustworthy sources were also used. Both quantitative and qualitative analysis techniques were used to achieve the research objectives.

Descriptive statistic's techniques such as frequency, mean, percentage and standard deviation were used to symbolize comprehensive information about the sampled households' socio-economic characteristics. Furthermore, different charts were used to present data illustratively.

Multiple regressions were used to identify the factors influencing the equivalent percentage of ICT and Non-ICT sources accessed. The qualitative data analysis was executed on the basis of the information collected from key informants' interviews, focus group discussion and field observations to make logical arguments and to draw appropriate conclusions.

2.4. Justification of variables

Two types of sources were accessed by the sampled farmers:

ICT Sources: Eight ICT sources, particularly Television, Radio, Mobile Phones, Landline Phones, Community Loudspeakers, Computer, Internet and Newspaper.

(a). Percentage Equivalent of ICT sources accessed out of Total ICT sources:

$= ((\sum S_iF_i) / SF) *100$

where,

S_i = ICT source i

F_i (Frequency of ICT source i) can be:

0 – None

1 – Yearly

2 – Seasonal

3 – Monthly

4 – Fortnight

5 – Weekly

6 – Daily

S = Total ICT sources

F = Maximum frequency = 6 (daily)

Non-ICT Sources: Eight Non-ICT sources, particularly KVKs, Public Extension Agent, Input Dealers, Output Dealers, Private Consultants, Other Farmers, Relatives, Friends, and Others.

(b). Percentage Equivalent of Non-ICT sources accessed out of total Non-ICT sources:

$= ((\sum S_iF_i) / SF) *100$

where,

S_i = Non-ICT source i

F_i (Frequency of Non-ICT source i) can be:

0 – None

1 – Yearly

2 – Seasonal

3 – Monthly

4 – Fortnight

5 – Weekly

6 – Daily

S = Total Non-ICT sources

F = Maximum frequency = 6 (daily)

(c). Percentage Equivalent of total ICT sources accessed out of total (ICT + Non-ICT sources)

= (Percentage Equivalent of ICT sources accessed out of total ICT sources) / 2

(d). Percentage Equivalent of total Non-ICT sources accessed out of total (ICT + Non-ICT sources)

= (Percentage Equivalent of Non-ICT sources accessed out of total Non-ICT sources) / 2

Note: As there were equal numbers of ICT and Non-ICT sources so both expressions (a) and (b) were divided by 2 to calculate (c) and (d).

(e). Percentage Equivalent of information through personal knowledge and experience= 100 – ((Percentage Equivalent of total ICT sources accessed out of total (ICT + Non-ICT sources) + (Percentage Equivalent of total Non-ICT sources accessed out of total (ICT + Non-ICT sources))

To evaluate the factors allied with the Percentage Equivalent of ICT and Non-ICT sources accessed as dependent variables (Y), the following multiple regression models have been achieved.

$$Y= \alpha+ \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6 + \beta_7X_7 + U$$

where, α constant term; β is the regression coefficients; X_i the socioeconomic factors; U the random disturbance term.

The explanatory variables (X_i) used in this analysis comprised:

X_1 – Gender

X_2 – Education Level

X_3 – Family Labor

X_4 – Years of farming experience/ age (years)

X_5 – Land holding size (acres)

X_6 – Off - Farm income (Rs.)

X_7 – Employment type

3. Empirical Results and Discussion

The percentage of male and female in the farm households were 95% and 5% respectively. Closely equivalent proportions of male and female were stated by Degu et al. (2015) in non-participant case; Babu et al. (2011).Nearly 94% respondents were married and the seven women in the sampled region were widows. Nearly identical results were given by Babu et al. (2011). Almost the same percentage of widow respondents was reported by Adefalu et al. (2013).

Of the sampled households, the secondary level education was the highest at 29%. Nearly same percentage equivalent of secondary level users was observed in a study by Gandhi (2014) whereas farmers with the Masters' degree or higher education - was the lowest at 2%.Also, many sampled farmers have had formal education. This manifests high literacy rate in the sampled region. This result is in conformity with of the Naveed and Anwar (2013).

Almost 73% sampled farmers were aged 34-59 years. The results of existing study are in coordination with the Rehman et al. (2013); Demiryurek et al. (2008); Omobolanle (2008); Ofuoku et al. (2008); Fawole (2006) who observed that most of the sampled farmers were in the middle-age category.

Majority of the respondents (58%) had 21-39 years of farming experience, and nearly 19% farmers had 40-58 years of farming experience. Most of the experienced farmers were involved in the advancement of innovation.

A minority of sampled farmers had land holding sizes greater than 25 acres. Analogous results by Naveed and Anwar (2013). Most farmers had low off-farm income, i.e., less than Rs.40,000 per year. Most of the sampled respondents (69%) work on a part-time basis in agriculture to earn the livelihood.

3.1. Association between Education Level and Employment Type of Sampled Farmers

The findings of the above test reveal that χ^2 (5) =24.557, p = 0.000. On the basis of this, it can be interpreted that there is a statistically significant association between education level and employment type; when the respondents attain higher educational, they prefer agriculture as a part-time source of income (Appendix A1).

Table 1 Results of chi-square tests and symmetric measures between education level and employment type

		Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square		24.557	5	0.000
Nominal by Nominal	Phi	0.430		0.000
	Cramer's V	0.430		0.000

Source: Computed from field survey, 2016

From the Phi and Cramer's V test, it can be concluded that the strength of association between variables is very high.

3.2. Influence of Education Level on Adoption of ICT

From the one-way analysis of variance (ANOVA), designated percentage equivalent of ICT sources accessed (Out of total ICT + Non-ICT) as a dependent variable and level of education as a fixed factor, it was found that adoption of ICT sources was higher for educated farmers (Appendix A2 and A3).

The analogous influence of education level was notified by Feder et al. (1985); Moghaddam and Abadi (2013); and Senthilkumar et al. (2013). The frequency and usefulness of information services of ICT projects in India, viz., Gyandoot and Warana are correlated to the farmers' education as indicated in Meera et al. (2004).

One more inference can be drawn by taking percentage equivalent of ICT sources accessed (Out of Total ICT + Non-ICT) as a dependent variable and education level as an independent variable in linear regression that is 64.2% variation in the dependent variable can be elucidated by education level of sampled farmers (Adjusted $R^2 = 0.642$).

3.3. Multicollinearity Conditions for Multiple Regression model of % equivalent of ICT sources accessed and % equivalent of Non-ICT sources accessed

Out of eight independent variables, two variables, i.e., age (years) and years of farming experience had tolerance value less than 0.1 and VIF>10. The multicollinearity statistics of age and years of farming experience is indicated in Table 2.

Table 2 Multicollinearity statistics for age and years of farming experience out of 8 independent variables

Variables	Collinearity Statistics	
	Tolerance	VIF
Age (Years)	0.069	14.506
Years of farming Experience	0.060	16.736

Source: Computed from field survey, 2016

For an independent variable's age, Tolerance = 0.069, i.e., (1-0.069) = 0.931 or 93.1% variation in age is explained by 7 other independent variables by considering them as a dependent variable.

Similarly, for years of farming experience, (1-0.060) = 0.94 or 94% variation in years of farming experience as a dependent variable can be explained by 7other independent variables including age.

To reduce multicollinearity effect, it is not possible to increase sample size; therefore, a new independent variable, i.e., years of farming experience/age (years) was used instead of using both age and years of farming experience separately. Adoption of this method shows no multicollinearity effect.

3.4. Utilizing ICT and Non-ICT Sources

For receiving agricultural information, the sampled farmers used ICT sources, Non-ICT sources and their own knowledge and experience.

Access to Agricultural Information by using of personal Knowledge and Experience

From Figure 2, it can be deduced that there is a phenomenal decrease in the percentage equivalent of information received through personal knowledge and experience on increasing equivalent rank of sampled farmers, i.e., rank from low to high ICT access.

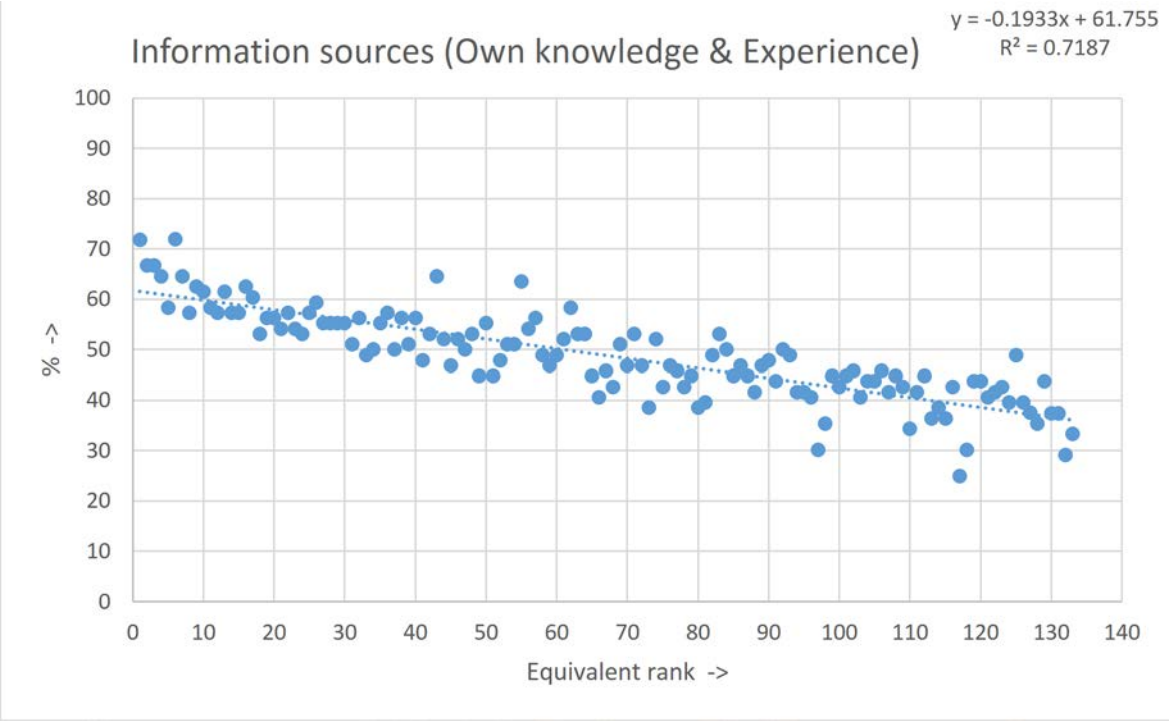


Figure 2 Percentage equivalent of information through own knowledge and experience vs. equivalent ICT rank

Factors influencing percentage equivalent of ICT sources accessed (Out of total ICT+Non-ICT)

Table 3 shows that the adjusted R^2 value of the fitted regression equation was 0.692, which indicated that 69.2 % of the variation in the percentage equivalent of ICT sources accessed (Out of total ICT + Non-ICT sources accessed) can be explained by the 7 independent variables used within the analysis.

The table further shows that out of seven independent variables taken as predecessors to the percentage equivalent of ICT sources accessed, four variables- gender, education, family labor and land holding size (acres) - were significantly linked to the percentage equivalent of ICT sources accessed at 0.05 level of probability. Out of 4 variables, the regression coefficients of two variables are positively allied to the percentage equivalent of ICT sources accessed displaying the direct relationship between education level and land holding size (acres) with the percentage equivalent of ICT sources accessed. The regression coefficients of other two variables, i.e., gender and family labor, were negatively interrelated to the percentage equivalent to ICT sources accessed.

The variable coefficients reveal that education level (0.744) is the most important factor impacting the percentage equivalent to ICT sources accessed. It is followed by gender (-0.215), family labor (-0.169) and land holding size (acres) (0.096).

The multiple regression results clearly state that the percentage equivalent of ICT sources accessed, i.e., adoption of ICT higher among the sampled farmers with higher education and bigger land holdings (acres). The homogeneous results were reported by Putler and Zilberman (1988); Batte et al. (1990); Bonny (1992); Gibbon and Warren (1992); Warren et al. (2000); Senthilkumar et al. (2013). The comparable positive influence of the farm size on adoption of modern ICT techniques was detailed by Mittal and Mehar (2013).

Table 3 Factors influencing percentage equivalent of ICT sources accessed

Variables (X_i)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	Constant (α)	27.16	3.467		7.835	0.000
X_1	Gender	-7.420	1.788	-0.215	-4.151	0.000*
X_2	Education Level	3.406	0.289	0.744	11.766	0.000*
X_3	Family Labor	-0.763	0.266	-0.169	-2.872	0.005*
X_4	Years of farming experience/age(years)	5.462	3.957	0.086	1.380	0.170
X_5	Land holding size (acres)	0.067	0.036	0.096	1.886	0.062*
X_6	Off- farm income (Rs.)	3.369E-6	0.000	0.007	0.101	0.920
X_7	Employment type	-0.888	1.179	-0.053	-0.753	0.453
	R^2		0.708			
	Adjusted R^2		0.692			
	F		43.282			
	N		133			

Dependent Variable: Percentage equivalent to ICT sources accessed (Out of total ICT+Non- ICT)

* Significant at 5% level ($p < 0.05$)

Ali and Kumar (2010) analyzed the impact of e-Choupal, an initiative of Indian Tobacco Company (ITC), and found that education, income and landholding size are important positive factors that influence the use of ICTs decision making. However, it is not clear how age factor impacts ICT adoption. Similar results were observed by Putler and Zilberman (1988); Gibbon and Warren (1992); Warren et al. (1996); Meera et al. (ikisan project) (2004). The percentage equivalent of ICT was found to be less among female farmers, and family with more household labors. The similar negative influence of gender was stated by Moghaddam and Abadi (2013); Senthilkumar et al. (2013).

Factors influencing percentage equivalent of Non-ICT sources accessed

It can observed from the Table 4 that the adjusted R^2 value of the fitted regression equation was 0.092, which indicated 9.2 % variation in the percentage equivalent of Non-ICT sources adopted can be elucidated by the 7 independent variables included in the analysis.

The table further unveils that out of 7 independent variables taken as predecessors to the percentage equivalent of Non-ICT sources accessed, two variables, i.e., education level and land holding size (acres) were significantly linked with the percentage equivalent of Non-ICT sources at 0.05 level of probability. Out of 2 variables, the regression coefficient of education level was negatively associated with the percentage equivalent of Non-ICT sources accessed displaying the indirect relationship of education level with percentage equivalent of Non-ICT sources accessed. The regression coefficient of land holding size (acres) was positively interrelated with the percentage equivalent of Non-ICT sources accessed.

The standardized beta coefficients undoubtedly revealed that education level (-0.258) was the most important factor impacting the percentage equivalent of Non-ICT sources accessed followed by land holding size (acres) (0.200). The multiple regression results clearly designate that as the sampled farmers had more land holding size (acres), the percentage equivalent of Non-ICT sources accessed was found to be more. This percentage equivalent of Non-ICT sources was contemplated to be less in case of educated farmers which showed more centralization of such farmers on information through ICT or own knowledge and experience.

The overall negative influence of significant factor was pointed out to be more compare to the positive influence.

Table 4 Factors influencing percentage equivalent of Non-ICT sources adopted

Variables (X_i)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	Constant (α)	31.602	3.517		8.985	0.000
X_1	Gender	-1.458	1.813	-0.072	-0.804	0.423
X_2	Education Level	-0.699	0.294	-0.258	-2.381	0.019*
X_3	Family Labor	-0.437	0.269	-0.163	-1.621	0.107
X_4	Years of farming experience/age(years)	-0.894	4.015	-0.024	-0.223	0.824
X_5	Land holding size (acres)	0.083	0.036	0.200	2.286	0.024*
X_6	Off- farm income (Rs.)	-3.536E-5	0.000	-0.132	-1.044	0.299
X_7	Employment type	-1.597	1.196	-0.162	-1.336	0.184
	R^2		0.141			
	Adjusted R^2		0.092			
	F		2.921			
	N		133			

Dependent Variable: Percentage equivalent of total ICT and Non-ICT sources accessed

* Significant at 5% level ($p < 0.05$)

4. Conclusions and Recommendations

Majority of the sampled farmers in the study area were male and married. Most sampled farmers had completed senior-secondary education. Only a few farmers had earned masters or other higher education. Most of the farmers were middle aged. The average family labor was contemplated to be three. Majority of the farmers had 2-6 family members and had fairly long experience in farming. Only a handful of sampled farmers had land holding size greater than 25 acres. Nearly for all farmers, farm income and off-farm income was less than Rs.40, 000 per year separately. Majority of the sampled respondents adopted agriculture as a part-time job to earn their livelihood.

The farmers with higher-education work in the agriculture on a part-time basis and while farmers with a low level of education require better productivity to increase their income in the bestowed research region. There was a positive correlation between education level and ICT adoption, and the major variation in percentage equivalent of ICT sources adopted can be explained by education of sampled farmers. The source of information (own knowledge and experience) was negatively related to equivalent ICT rank of sampled farmers. The education and land holding size (acres) were positively correlated with the percentage equivalent of ICT sources accessed whereas household labor was negatively impacting it. The percentage equivalent of ICT sources accessed was less in case of female farmers. The most influencing factor is education of sampled respondents.

The prominent factors influencing percentage equivalent of Non-ICT sources accessed were education level and land holding size (acres). The educated farmers were more focused on information through own knowledge and experience or ICT sources. The education level was the most influential negative factor and land holding size (acres) had positive influence on the percentage equivalent of Non-ICT sources accessed. The government intervention is mandatory to co-ordinate and regulate the efforts of the private sector as ICT bears infrastructural, developmental and expenditures to ensure mobile networks connect the farmers to information so that highly educated farmers with proper ICT

access can adopt agriculture as a full-time profession and ensure the growth of agricultural mechanization. The influences of both (ICT + Non-ICT) sources and information through personal knowledge and experiences can be used in the research for comparative analysis.

References

Adefalu, LL, Aderinoye-Abdulwahab, SA, Bello, OG, Olorunfemi, OD, and Oba, SA 2013, 'Information needs of fish farmers in Ilorin metropolis', Kwara state, Nigeria. Nigerian Journal of Agriculture, Food and Environment, 9(2), pp. 1-5.

Ali, J 2012 'Factors affecting the adoption of Information and Communication Technologies (ICTs) for farming decisions', Journal of Agricultural & Food Information, Vol. 13, Issue 01, pp. 78-96. https://doi.org/10.1080/10496505.2012.636980

Ali, J & Kumar, S 2010, 'Information and communication technologies (ICTs) and farmers' decision-making across the agricultural supply chain', International Journal of Information Management, 31 (2011), pp. 149–159. https://doi.org/10.1016/j.ijinfomgt.2010.07.008

Asian Development Bank 2004, 'Building e-community centers for rural development', Report of the regional workshop, Bali, Indonesia.

Babu, SC, Glendenning, CJ, Asenso-Okyere, K & Govindarajan, SK 2011, 'Farmers' information needs and search behaviors: Case study in Tamil Nadu, India', International Food Policy Research Institute, Final draft.

Batte, MT, Jones E & Schnitkey GD 1990, 'Computer use by Ohio commercial farmers', American Journal of Agricultural Economics, Vol. 72, No. 4, pp. 935-945. https://doi.org/10.2307/1242625

Bonny, S 1992, 'Ongoing technical change on farm holdings in a developed country, France. A survey on its vectors, its nature and the farmers concerned.', Agricultural Systems, Vol. 38, Issue 1, pp. 75-103. https://doi.org/10.1016/0308-521X(92)90088-6

Degu, G, Alemayehu, F, Mekonen, S & Abachew, D 2015, 'Socioeconomic characteristics of common bean variety adoption and seed quality survey in southern region, Ethiopia', International Journal of Scientific & Technology Research, Vol. 4, Issue 07, pp. 234-237.

Demiryurek, K, Erdem, H, Ceyhan, V, Atasever, S & Mayıs, OUO 2008, 'Agricultural information systems and communication networks: the case of dairy farmers in the Samsun province of Turkey', Info. Res. [Online], 13(2). Available at: http://dlc.dlib.indiana.edu/dlc/bitstream/handle/10535/5731/agricultural%20information%20systems...pdf?sequence=1.

Dutta, R 2009, 'Information needs and information seeking-behavior in developing countries: A review of the research', International Information & Library Review, Vol. 41, no.1: 44-51. https://doi.org/10.1080/10572317.2009.10762796

Gelb, E, Maru, A, Brodgen, J, Dodsworth, E, Samii, R & Pesce, V 2008, 'Adoption of ICT Enabled Information Systems for Agricultural Development and Rural Viability', The Global Forum on Agricultural Research. Pre-conference Workshop Summary. August 2008.

Gibbon J & Warren MF 1992, 'Barriers to adoption of on-farm computers in England', Farm Management, 8(1), pp. 37-45.

Harris, RW 2004, 'Information and communication technologies for poverty alleviation. United Nations Development Program', Asia-Pacific Development Information Program (UNDP-APDIP).

Isife, BI, and Ofuoku, AU 2008, 'Communication in Agricultural Extension and Rural Development: Methods and Concepts' Owerri, Springfield Publication.

Just, DR, Wolf, SA, Wu, S & Zilberman, D 2002, 'Consumption of economic information in agriculture', American Journal of Agriculture Economics, Vol. 84, Issue 1, pp. 39–52. https://doi.org/10.1111/1467-8276.00241

Karamagi Akiiki, E 2006, 'Towards improving farmers'livelihoods through exchange of local agricultural content in rural Uganda', Knowledge Management for Development Journal 2(1):68- 77.

Koehnen, TL 2011, 'ICTs for Agricultural Extension: Global Experiments, Innovations and Experiences' The Journal of Agricultural Education and Extension, Vol. 17, Issue 5, pp. 473-474. https://doi.org/10.1080/1389224X.2011.624714

Kumar, G & Kumar, RS 2012, 'Impact of Information and Communication Technology in Agriculture–Perception of the Farmers in Ramanathapuram District', International Journal of Statistika and Mathematika, Vol. 4, Issue 2, pp. 33-41.

Malhan, IV & Rao, S 2007, 'Impact of Globalization and Emerging Information Communication Technologies on Agricultural Knowledge Transfer to Small Farmers in India', Department of Library Science. Paper presented in World Library and Information Congress: 73rd IFLA General Conference and Council, Durban, South Africa. 19-23 August 2007.

Martin, BL, Stewart, DL & Hillison, J 2001, 'Computer anxiety levels of Virginia Extension personnel', Journal of Extension, 39 (1), www.joe.org/joe/2001february/a1.html.

Meera, SN, Jhamtani, A & Rao, DUM 2004, 'Information and communication technology in agricultural development: A comparative analysis of three projects from India', Agricultural Research & Extension Network, Paper 135. Available from: http://www.odi.org.uk/sites/odi.org.uk/files/odi-assets/publications-opinion-files/5186.pdf.

Michiels, S & Van Crowder, L 2001, 'Discovering the "magic box": Local Appropriation of Information and Communication Technologies (ICTs) FAO, Rome', Retrieved July 14, 2004, from http://www.fao.org/sd/2001/KN060 2a_en.htm.

Mittal, S & Mehar, M 2013, 'Agricultural information networks, information needs and risk management strategies: A survey of farmers in Indo- Gangetic plains of India', Vol. 10.CIMMYT.

Moghaddam, BK & Abadi, AK 2013, 'Factors affecting ICT adoption among rural users: A case study of ICT center in Iran', Telecommunication policy, Vol. 37, Issue 11, pp. 1083-1094.

Naveed, MA & Anwar, MA 2013, 'Agricultural information needs of Pakistani farmers', Malaysian Journal of Library & Information Science, Vol. 18, no. 3, pp. 13-23.

Ofuoku, AU, Emah, GN & Itedjere, BE 2008, 'Information utilization among rural fish farmers in Central Agricultural Zone of Delta state, Nigeria', World J. Agric. Sci., 4 (5), pp. 558-564.

Omobolanle, OL 2008, 'Analysis of extension activities on farmers' productivity in Southwest, Nigeria', African Journal of Agricultural Research Vol. 3 (6), pp. 469-476.

Park, DB, Cho, YB & Lee, M 2007, 'The use of an e-learning system for agricultural extension: A case study of the Rural Development Administration, Korea', The Journal of Agricultural Education and Extension, Vol. 13, Issue 4, pp. 273-285. https://doi.org/10.1080/13892240701631000

Pereira, JV 2009, 'The new supply chain's frontier: Information management', International Journal of Information Management, Vol. 29, Issue 5, pp. 372-379. https://doi.org/10.1016/j.ijinfomgt.2009.02.001

Pokhrel, DM & Thapa, GB 2007, 'Are marketing intermediaries exploiting mountain farmers in Nepal? A study based on market price, marketing margin and income distribution analyses', Agricultural Systems, Vol. 94, Issue 2, pp.151-164. https://doi.org/10.1016/j.agsy.2006.08.004

Putler, DS & Zilberman, D 1988, 'Computer use in agriculture: evidence from Tulare County, California', American Journal of Agricultural Economics, Vol. 70, No. 4, pp. 790-802. https://doi.org/10.2307/1241920

Rees, DJ, Momanyi, M, Wekundah J, Ndungu, F, Odondi, J, Oyure, AO, Andima, D, Kamau, M, Ndubi, J, Musembi, F, Mwaura L & Joldersma R 2000, 'Agricultural knowledge and information systems in Kenya: implications for technology dissemination and development', Agricultural Research and Extension Network (AgREN) Network Paper no. 107. London: Overseas Development Institute.

Rehman, F, Muhammad, S, Ashraf, I, Mahmood, K, Ruby, T & Bibi, I 2013, 'Effects of farmers' socioeconomic characteristics on access to agricultural information: Empirical evidence from Pakistan', The journal of animal & plant sciences, 23(1): pp. 324-329.

Senthilkumar, S, Chander, M, Pandian, ASS & Kumar, NKS 2013, 'Factors associated with utilization of ICT enabled Village Information Centers by the dairy farmers in India: The case of Tamil Nadu', Computers and Electronics in Agriculture, Vol. 98, pp. 81-84. https://doi.org/10.1016/j.compag.2013.07.018

Stefano, LA, Stilwell, C, Morris, C, Hendriks, SL, Miya, D, Mkhize, T, Makhanya, T, Makhanya M, Ndlovu G & Sabelo D 2005, 'An action research study of the agricultural knowledge and information systems of small-scale commercial organic farmers in Umbumbulu, KwaZulu-Natal', Food Security Programe Project Brief no. 2. Pietermaritzburg: University of KwaZulu-Natal. 6 April 2007.

Warren MF, Soffe RJ& Stone MA 1996, 'The uptake of new communication technologies in farm management: A case study from the South West of England', Farm Management, 9 (7), pp. 345 – 356.

Warren MF, Soffe RJ, and Stone MAH 2000, 'Farmers, computers and the internet: a study of adoption in contrasting regions of England', Farm Management, Vol.10, No.11, pp.665-684.

APPENDIX A

Table A1 Cross-tabulation of education Level * employment type

			Employment type		Total
			Part	Full time	
Education Level	Illiterate	Count	8	18	26
		% within Education Level	30.8%	69.2%	100.0%
		% within Employment	8.7%	43.9%	19.5%
		% of Total	6.0%	13.5%	19.5%
	Primary	Count	22	8	30
		% within Education Level	73.3%	26.7%	100.0%
		% within Employment	23.9%	19.5%	22.6%
		% of Total	16.5%	6.0%	22.6%
	Secondary	Count	29	10	39
		% within Education Level	74.4%	25.6%	100.0%
		% within Employment	31.5%	24.4%	29.3%
		% of Total	21.8%	7.5%	29.3%
	Senior Secondary	Count	13	2	15
		% within Education Level	86.7%	13.3%	100.0%
		% within Employment	14.1%	4.9%	11.3%
		% of Total	9.8%	1.5%	11.3%
	Bachelor	Count	17	3	20
		% within Education Level	85.0%	15.0%	100.0%
		% within Employment	18.5%	7.3%	15.0%
		% of Total	12.8%	2.3%	15.0%
	Master or higher	Count	3	0	3
		% within Education Level	100.0%	0.0%	100.0%
		% within Employment	3.3%	0.0%	2.3%
		% of Total	2.3%	0.0%	2.3%
Total		Count	92	41	133
		% within Education Level	69.2%	30.8%	100.0%
		% within Employment	100.0%	100.0%	100.0%
		% of Total	69.2%	30.8%	100.0%

Table A2 ANOVA for % equivalent of ICT sources accessed (Out of total ICT+ Non-ICT) (dependent variable) and education level (fixed factor)

Source	Sum of Squares	df	Mean Square	F	Sig.
Education level	6133.614	5	1226.723	88.764	0.000
Error	1755.142	127	13.820		
Corrected Total	7888.756	132			

Table A3 Descriptive statistics of Education level

Education Level	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Illiterate	16.230	0.729	14.787	17.673
Primary	21.254	0.679	19.911	22.597
Secondary	30.319	0.595	29.141	31.497
Senior Secondary	30.420	0.960	28.521	32.319
Bachelor	34.587	0.831	32.942	36.232
Master or higher	38.893	2.146	34.646	43.141

Dependent Variable: % equivalent of ICT sources accessed (Out of total ICT+ Non-ICT)

APPENDIX B

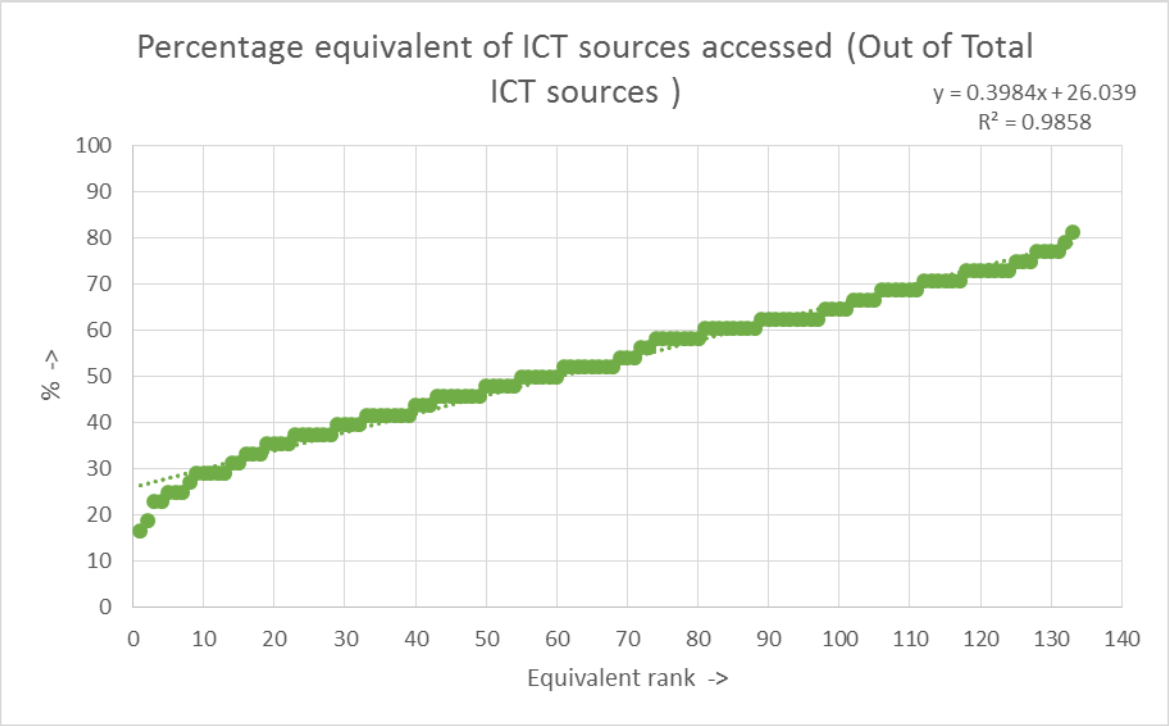


Figure B1 Percentage equivalent of ICT sources accessed (Out of total ICT sources) vs. equivalent rank

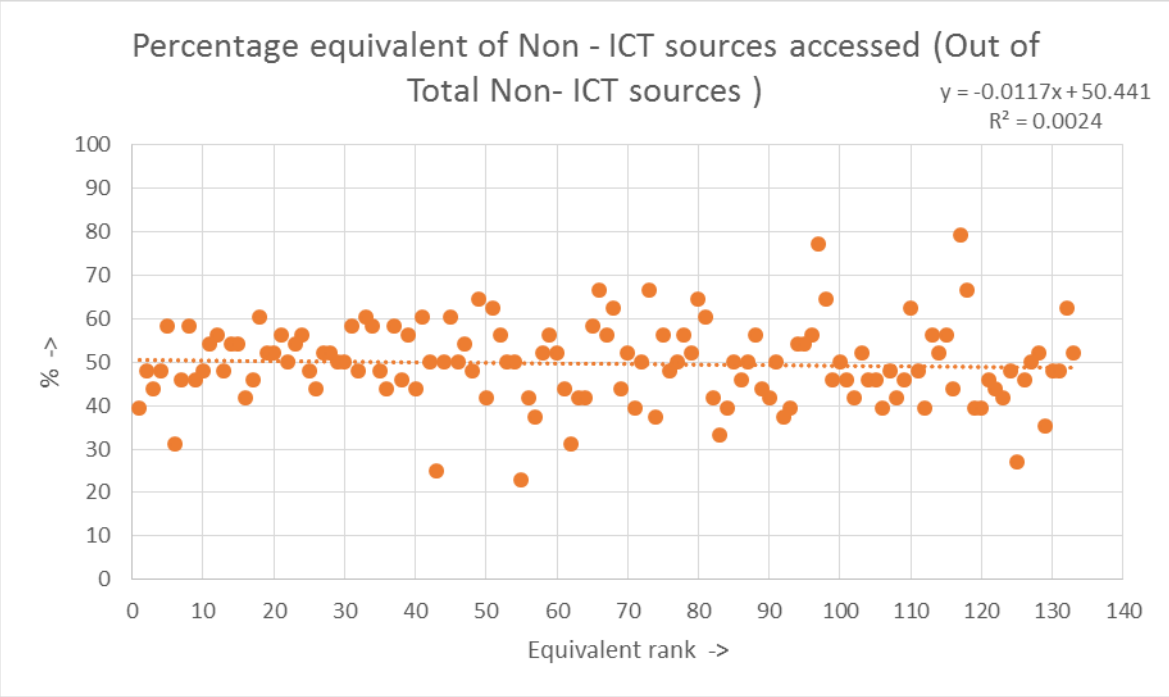


Figure B2 Percentage equivalent of Non-ICT sources accessed (Out of total Non-ICT sources) vs. equivalent rank

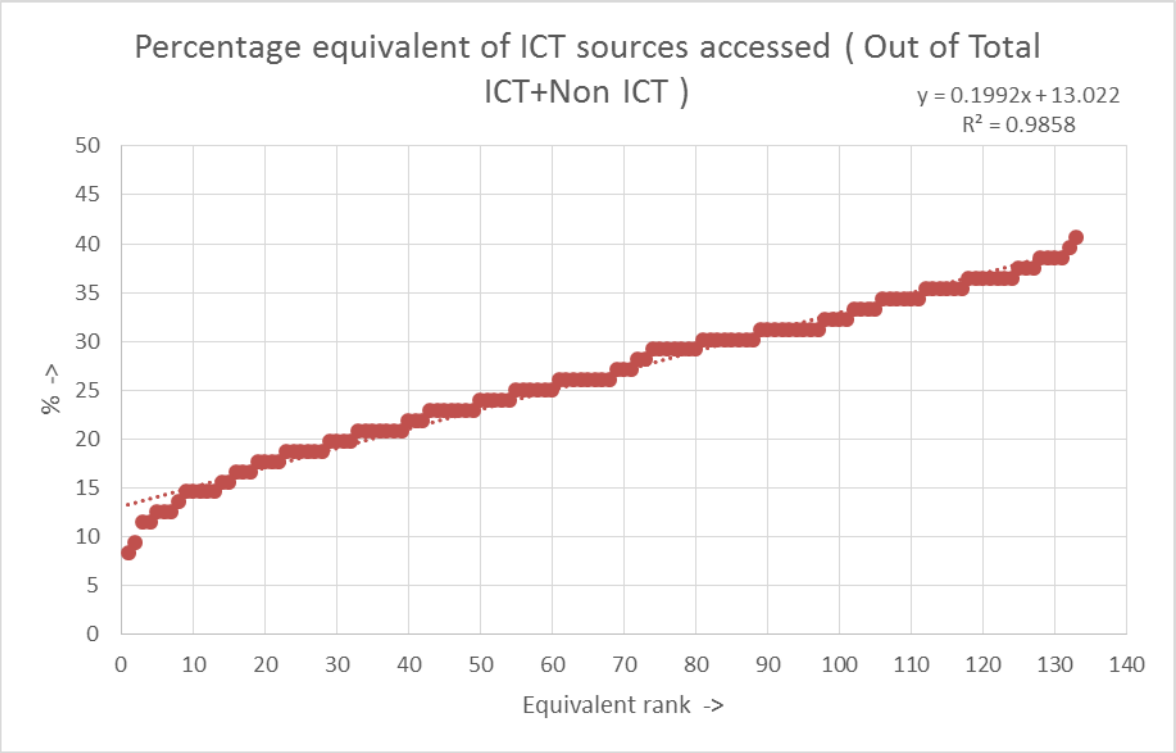


Figure B3 Percentage equivalent of ICT sources accessed (Out of total ICT + Non-ICT) vs. equivalent rank

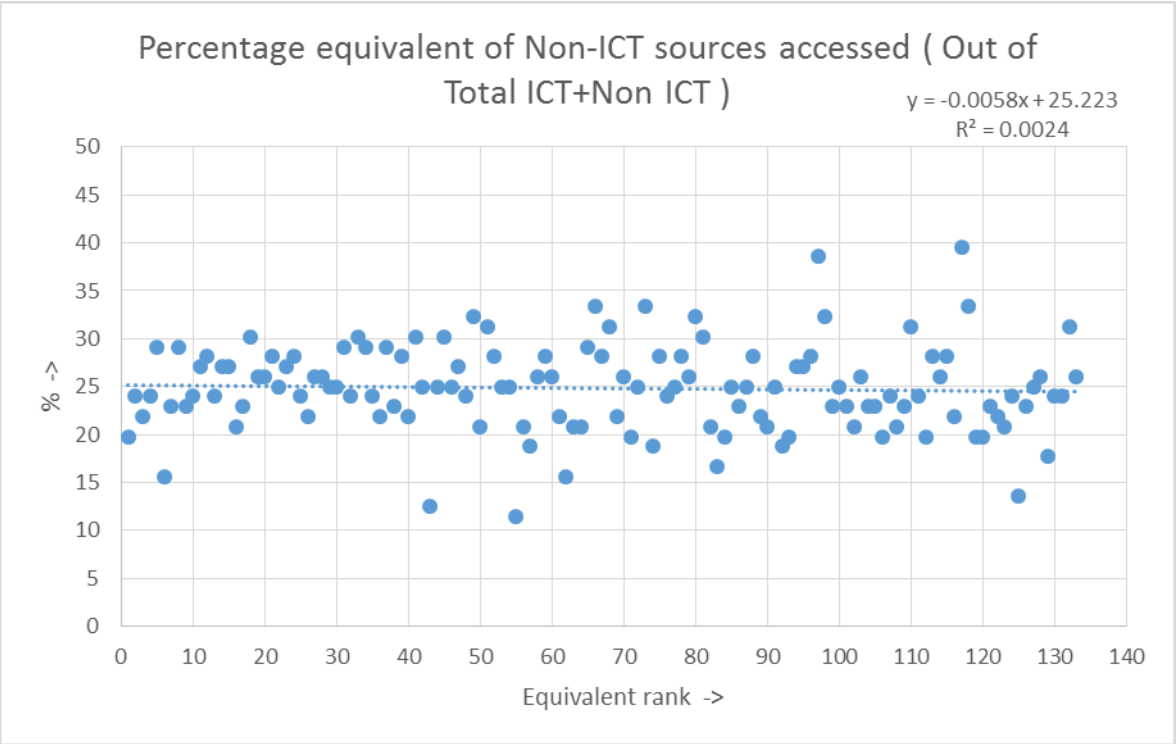


Figure B4 Percentage equivalent of Non-ICT sources accessed (Out of total ICT + Non-ICT) vs. equivalent rank

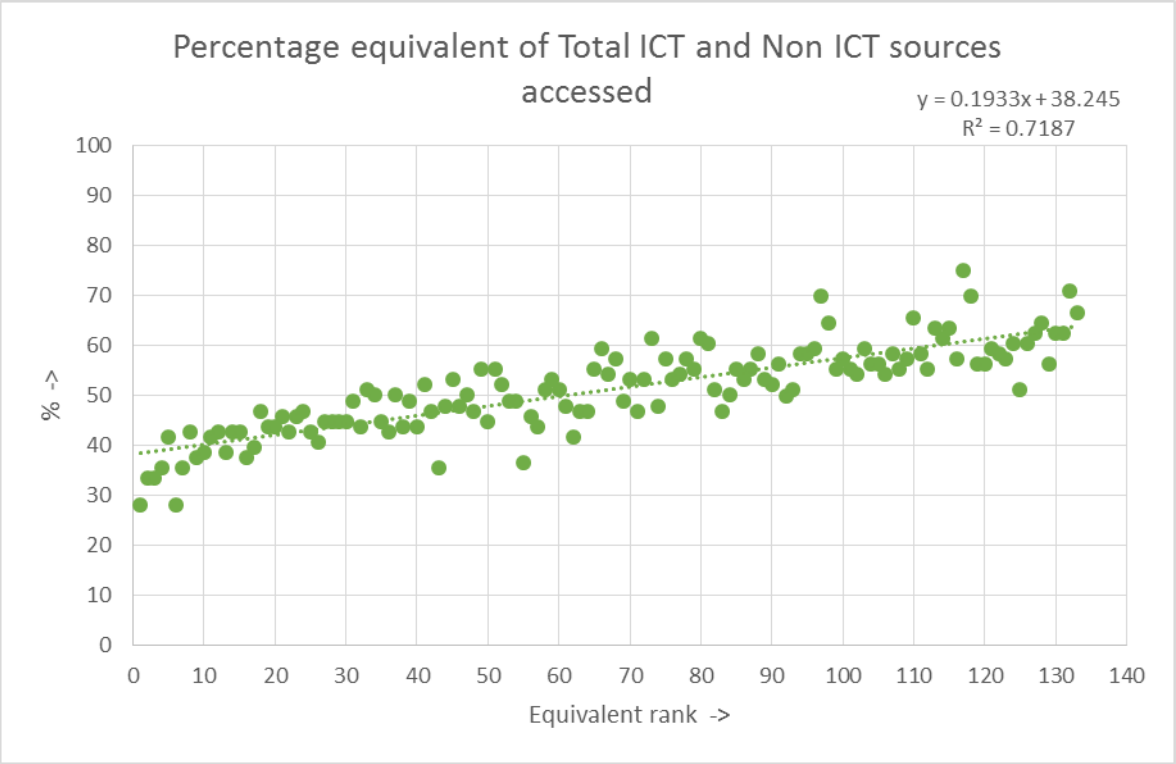


Figure B5 Percentage equivalent of total ICT and Non-ICT sources accessed vs. equivalent rank

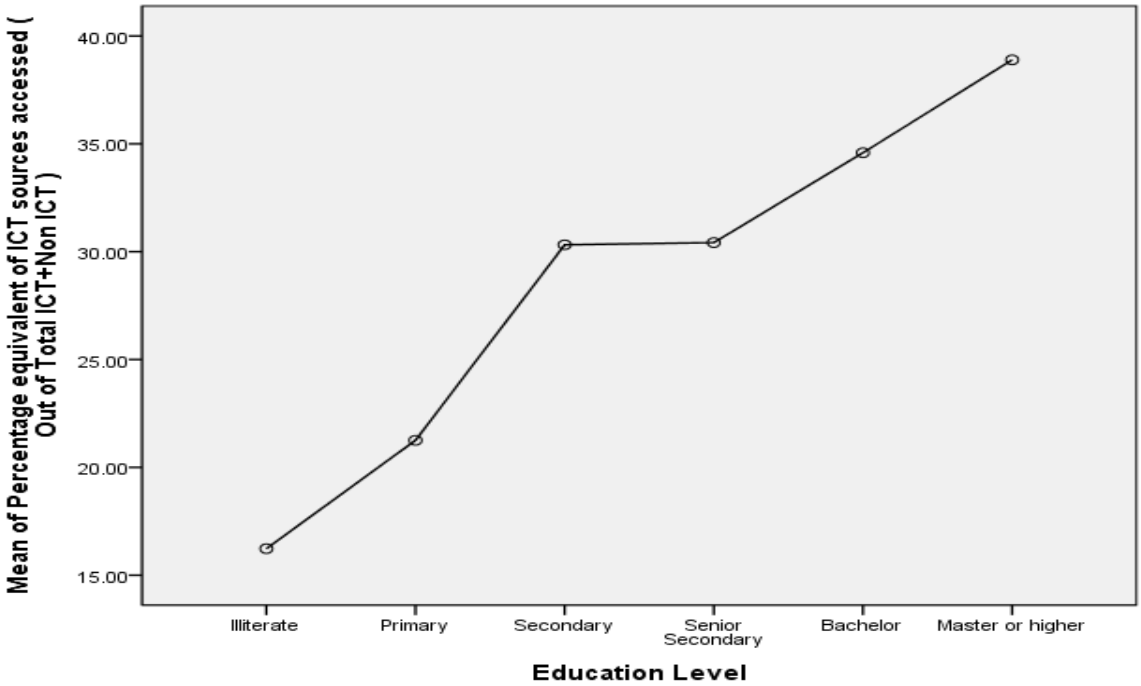


Figure B6 Mean of percentage equivalent of ICT sources accessed (Out of total ICT + Non-ICT) vs. education level

Hungarian Association of Agricultural Informatics
European Federation for Information Technology in Agriculture, Food and the Environment

Journal of Agricultural Informatics. Vol. 9, No. 2
journal.magisz.org

Overview of the electronic public services in agriculture in Jordan

Nemer Abu Rumman [1], Róbert Szilágyi [2]

I N F O
Received 11 Jan 2018
Accepted 16 May 2018
Available on-line 18 Jun 2018
Responsible Editor: M. Herdon

Keywords:
e-Government, agriculture, Jordan, electronic public services, evaluation of quality

A B S T R A C T

Currently, the use of information technology is essential and offering electronic services which has become a necessity for all governments around the world. In Jordan, ICT sector is advanced in the Middle East and provides electronic services in various sectors. The goals of this paper are to find out the current situation and evaluate the quality of electronic public services for the Jordanian agricultural sector. In this paper, the government websites providing electronic services were analyzed to check the readiness of these websites and identify and verify electronic services in the agricultural sector. Seven main sites have been selected concerned with the agricultural sector in Jordan. Only three websites offer electronic services and in order to check these websites a survey of these websites was conducted based on literature criteria. These criteria checked the readiness of these websites from technical aspects, services, information security and ease of use for people with special needs. After the portal analysis, these websites showed general weakness in the management of sites, weakness in security tools for personal information, weakness in updated information and no websites showed any features for people with special needs. In addition to it, farmers and agricultural businesses in Jordan are currently not obliged by any law to use any electronic service, and they can still opt for traditional paper mail or personal visits.

1. Introduction

Since the internet has become a vital tool for the success of businesses, it was very necessary for the governments to exploit its use and provide the best services for their stakeholders and citizens. The revolutionary emergence of information and communication that took place in 1980s and 1990s did not only change the lifeline of people, but also the whole characteristics of the interaction between the governments and their citizens. The public and private sectors in the mid-1990s competed on how best they exploit the new technologies in improving their services and developing their relationships with their stakeholders (Szilágyi & Szilágyi 2009). In general as many previous studies have shown (Gatautis R., Medziausiene A., Tarute A. n.d.; Seri & Zanfei, 2016; Córdova et al., 2014), the private sector showed more enthusiasm towards using ICT for communications, interactions and transactions; whereas the governmental sector was hesitant about using them in the beginning. Nonetheless, this hesitation did not last long and it quickly dissipated by the early 2000s when the governments started to launch e-Government projects. Most of the technology experts found that the spread of e-Government in 2003 was surprisingly better than how it was a decade before and the e-government term started being more familiar than any time before (United Nations 2002; Gil-García & Pardo 2005; Ebrahim & Irani 2005). However, since it became important to all government to convert to the electronic services governments all over the world started converting their services to electronic in all sectors and the agriculture sector is no exception, here this paper identifies the current situation for the electronic services in the agriculture sector in Jordan.

Finally, this paper (first stage) tries to explore the current situation of the e-Government in Jordan. The basic idea was that the e-government has positive effects in terms of reducing costs, reducing bureaucracy and corruption in a country with many economic problems and challenges. On the other

1 Nemer Abu Rumman
University of Debrecen
nemer.aburumman@econ.unideb.hu
2 Róbert Szilágyi
University of Debrecen
szilagyi.robert@econ.unideb.hu

hand this paper will explore the current electronic services in the agriculture sector, in particular evaluating the current governmental websites in agricultural sector from the technical aspect, and counting the number of electronic and non-electronic services. Finally the quality of these websites were analyzed and some recommendations and suggestions were offered to enhance it.

In addition, because the agricultural sector is considered as one of the most important sectors that contribute to the economy and at the same time the importance of technology in improving this sector, this paper is about the current situation of the electronic services for the agricultural sector in Jordan and the services will be examined further and compared with other countries in other stages later.

1.1. E-Government

There are a numerous number of studies that attempted to define e-Government. They all converged on the idea that e-Government is the government's use of technology that aims to better serve its stakeholders. For instance, the World Bank defines it as "the governments' use of information technologies, such as the Wide Area Networks, the internet, and Mobile Computing, which have the potential to transform the traditional paper-based relationship between businesses, citizens, and governments to a more modern one. These technologies are thought to have a variety of good consequences, like better service delivery to citizens". E-Government is also "the use of ICT to improve, transform, and redefine any form of information exchange among the involved actors by developing and maintaining the dedicated inter-organizational systems, virtual organizational arrangements, and (inter)national institutional arrangements." (Rahim & Athmay 2013). Similarly, e-Government refers to the administration system in the government offices that uses modern technologies, like information, network and office automation technologies to handle the official affairs and provide public services for society (Lin et al. 2011).

Irrespective of having different definitions about e-Government, there is one common central concept that underlies all the above-mentioned definitions which is the use of technology or (ICT) by government' agenesis for public service delivery through different platforms and channels.

1.2. Electronic Services in agriculture (E-agriculture)

Agricultural sector is one of the most important sectors in the world because it is the main Food supplier. The technology in agriculture sector is still important and a national and global priority (Gelb & Parker 2005), but according to a study conducted by (Karetsos 2014), the agricultural sector is still one of the business sectors that have been left aside in terms of the application of new technologies, so agricultural information system needs to be developed and increase the electronic services.

ICTs are one of the main contributors to growth and socio-economic development in business sectors, countries and regions where they are well dealt with. The good adoption and integration of ICTs have improved service delivery, created new jobs and opportunities, saved time and money (Botos et al. 2015). Moreover, many previous studies have shown more benefits like: promoting greater inclusion in the broader economy, raising efficiency by complementing other production factors, and fostering innovation by dramatically reducing transaction costs (Deichmann et al. 2016). Meanwhile, the rapid growth of mobile phone globally (Figure 1) provides new possibilities to access and share information. About more than half of the population in our world owns a mobile phone and it would be much higher when children are not counted.

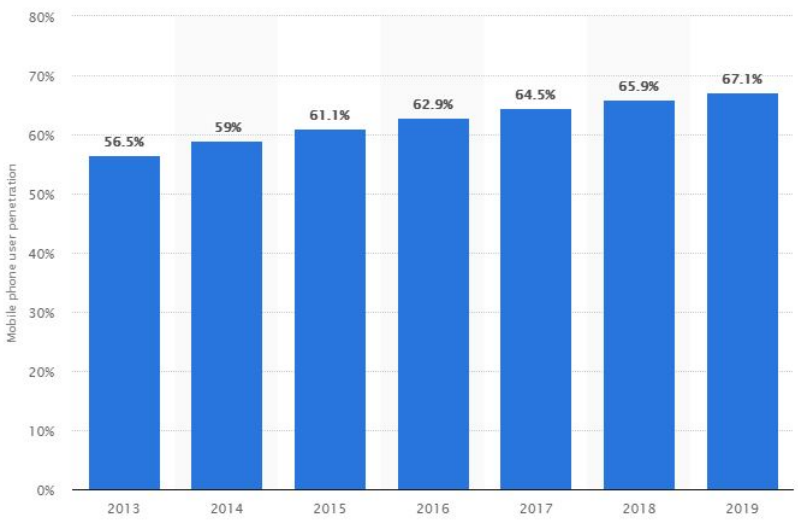
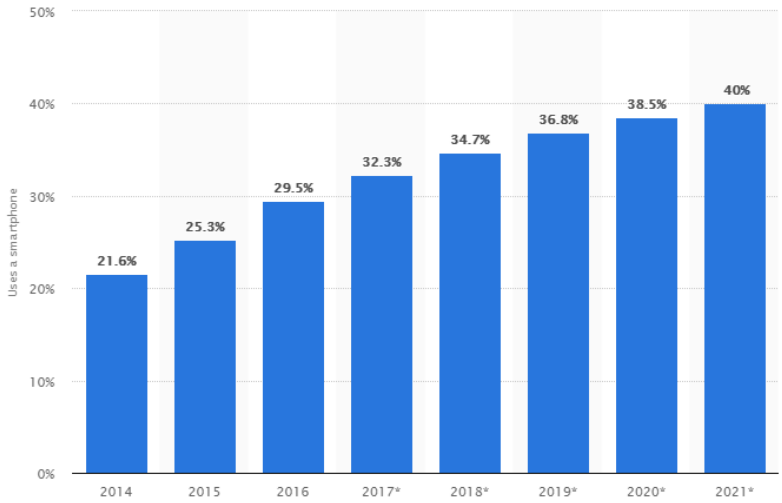


Figure 1. Mobile phone user penetration as percentage of the population worldwide from 2013 to 2019 (Statista 2016)

But in the other hand, many farming communities many people with old phones still rely on basic phones, which offer voice and text services mainly, but smart phone access is on the rise (Figure 2). The rapid growth of mobile broadband (Statista 2016) can provide a great opportunity for e-agriculture (Karetsos 2014). In sum, as second edition of the FAO report mentioned that ICTs have provided new opportunities to address the challenges faced by agriculture as well as information and knowledge gaps that ICT can also play a very important role in bridging these gaps in the sector. (Lengyel et al. 2015; FAO 2017)



).

Figure 2. Smartphone user penetration as percentage of total global population from 2014 to 2021 (Statista 2016)

2. E-Government in Jordan

The Jordanian ICT sector has grown tremendously since the late 1980s and early 1990s and since the establishment of Ministry of Information and Communications Technology (MOICT) in 2002, the ministry became more involved and enhanced cooperation within the ICT sector (MoICT 2016). Today, with about 89% of the Jordanian populations connected to the internet (We Are Social 2017). ICT sector is considered as one of the main sectors of the Jordanian economy With a GDP contribution worth around 4 percent in 2017 and more 84,000 jobs in this sector the ICT sector has witnessed significant growth over the years (Jordan Times 2017) (Figure 3).

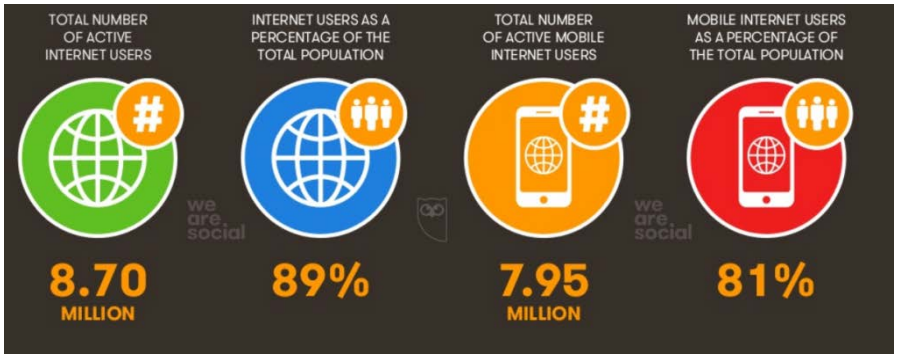


Figure 3. Jordanian internet and mobile phone users in 2017. (We Are Social 2017)

Jordan considers that e-Government implementation will provide new channels and means of access to information and data services that result in economic and social developments. Jordan began to pursue the project of e-Government in 2001 after King Abdullah II launched the e-Government initiative, and in April 2002 after establishing MOICT to be responsible for articulating policy in the areas of Information Technology, Telecommunications. The king put the ministry in charge of implementing the e-Government project throughout the country (MoICT 2016). The e-Government portal was launched in late 2006. The main aim of the government was to orientate itself towards the customer approach and providing services to people, regardless of their location, economic status, education, or ICT abilities.

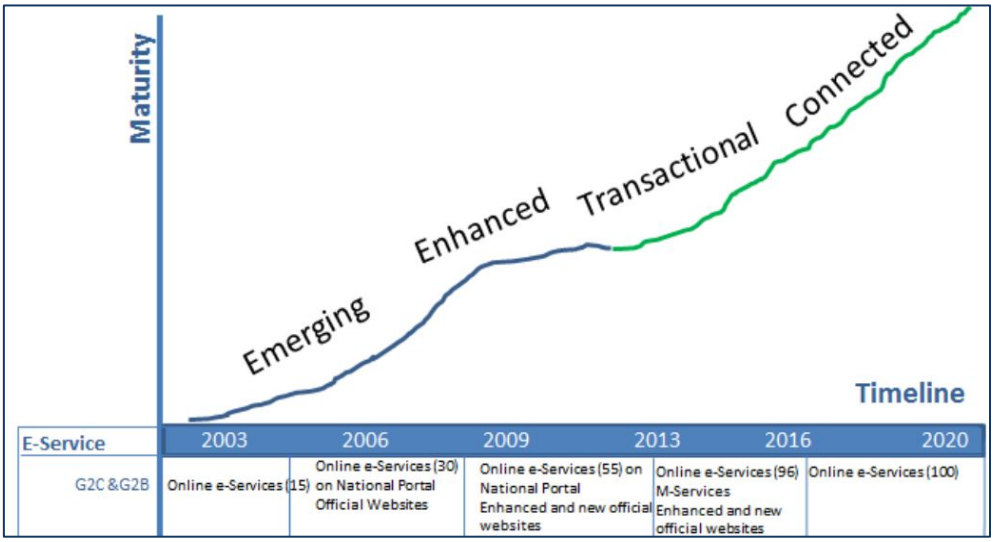


Figure 4. Jordan e-Government Current and Target Maturity (MoICT 2013)

According to the Ministry of Information and Communication Technology (MoICT) the Jordanian e-government milestones showed in Figure 4. The report showed that since the launch of the program in 2001, the program has achieved some successes and faced some challenges, during 2001-2003, Jordanian government and the MoICT reviewed the process to implement e-government in Jordanian society and major plan at that time was to set and develop roadmaps to guide the implementation of the e-government project. Based upon the priorities defined in the roadmap, they distinguished five governmental institutions to be the first institutions to apply e-services (MoICT 2016).

Afterwards, the program has produced a National e-Government Strategy and Roadmap (2006-2009) and during this time the program succeeded in the establishment of 107 Units of e-Government in government agencies, with active reporting mechanism between those agencies and e-Government Program as well (G2G) and Launch a Gateway for SMS that supports 77 government agencies (MoICT 2016).

In 2010 the government has produced a strategy for the years (2010-2013) and at the end of this strategy the program activated many services like "Jordanian Payment Gateway" which enables users

to make an online payment, "Mobile Portal" that provide 40 e-services to mobile customers. In addition to it, the program gave over 13000 training opportunities on ICT literacy, project management, vendor certified training programs and IT professionals at this stage. Currently Citizens and Businesses have access to about 100 e-Services on the National Government Portal provided by 45 governmental agencies include police clearance, drivers, and vehicle licensing, e-Tax, a real estate registry, and national exam registration and results. An additional 70 e-services will be launched in 2016, according to the ICT Minister in Jordan (MoICT 2016).

2.1. Agricultural Sector in Jordan

Agriculture in all countries is considered as one of the economic pillars. And recently, agriculture has started to become a major role in the protection of the environment and ensuring an environmental balance that would secure sustainable use of resources and preserve them for future generations. In Jordan, despite the sector's declining contribution to GDP in the last decades (4.0%) (Figure 5) and employs seven percent of the labor force according to the Jordanian Department of Statistic. But it is still considered the base for integrated rural development, a source of income and employment for rural and Badiya (desert) people and a generator of activities in the other economic sub-sectors.

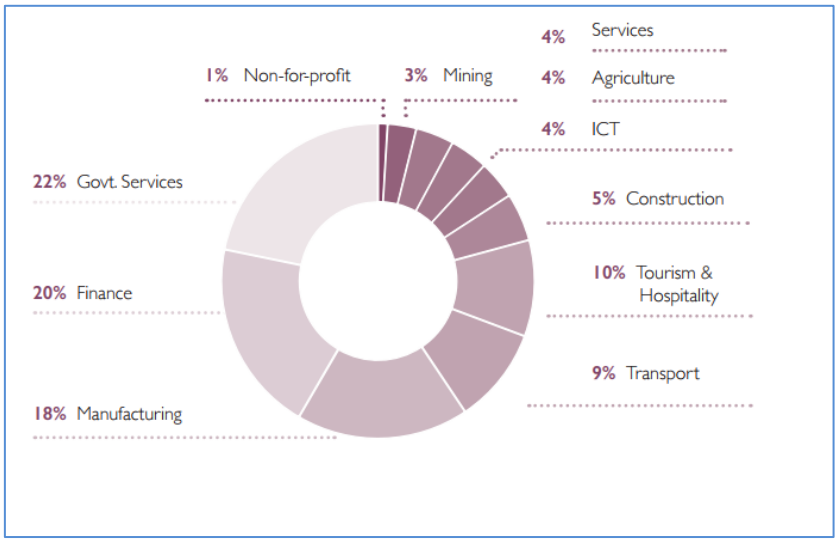


Figure 5. Sectoral Contribution to the Jordanian Economy (Department of Statistics 2017)

On other hand, the Sector is still facing a large number of challenges which is increased with the low and fluctuating rainfall, successive of drought years and environmental changes as well as the various risks related to the decline in the role of the agricultural sector at the national economy which leads to focus all the efforts to enhance agricultural development that connected with the recent and future plans for economic and social developments, political and environmental developments to reduce the negative effects and to promote its sustainability.

One of the most important factors that influences the improvement of the agricultural sector is the use of technology and that has shown impact in many countries around the world (FAO and International Telecommunication Union 2016) but in Jordan, the use of technology in the agricultural sector is still limited and substandard (Saravanan 2010). In this paper, the governmental websites and the electronic services of the agriculture sector were reviewed.

2.2. Electronic Services in agriculture sector in Jordan

Many governments' agencies around the world are already developed and have implemented portals and websites where various electronic services for farmers are provided (Ntaliani et al. 2010). After surveying the governmental websites it was revealed that Jordan offered agricultural services on the internet through three main websites.

First the Jordanian e-Government website (https://jordan.gov.jo), second the Jordanian Ministry of Agriculture (http://www.moa.gov.jo) and the National Agricultural Information System (NAIS) (http://www.nais-jordan.gov.jo). And four other websites:

1. National Center for Agricultural Research and Extension (NCARE). (http://www.ncare.gov.jo)
2. Bee Research Department. (http://www.jordanbru.info/)
3. Olive Research and Extension Program in Jordan. (http://www.ncare.gov.jo/OliveProgram)
4. Agricultural Credit Corporation. (http://www.acc.gov.jo)

However, it is noted that there is no unified gate that deals only with agricultural services or the agricultural sector. In this paper, we scanned all the agriculture related websites run by the government agensis for farmers, these websites were reviewed and nine technical criteria were used to determine the readiness of these websites to provide services.

3. Methods

The objective of this paper is to analyze the Front office of e-Government offering agricultural services according to (Budai & Tózsa 2011) "is the face of public administration and it includes all interfaces, where public administration and customer are in contact with each other according to this meaning, there are also here online and offline solutions". Although nowadays front office solutions mean practically electronic solutions in this interface where the user and the office or the administration can meet and the user can access to all the services related to him/her.

In addition, e-Government front office is electronic where the communication between the user and the office is electronic without physical presence or any time limit (24/7) for example:

- CRM-system,
- Official websites portals,
- mobile portals.

Here, in this paper we will focus on the second channel (The official websites portals) that provides electronic services in the agriculture sector and tries to assist them by checking the quality of these websites, the readiness of them to offer electronic services, how many services they provide and the quality of these services, here below in Table 1. It is shown the eight websites that are related to agriculture in Jordan. This table shows the agencies who manage these websites, if the information provided is useful, and the number of services provided, both in electronic or non electronic services.

Table 1. Websites related to agriculture in Jordan (own source)

Websites / criteria	Responsible party	Useful information	Number of services	Offline services	Online services
https://jordan.gov.jo	**Ministry of Information and Communication Technology**	**Yes**	**2037**	**1888**	**149**
http://moa.gov.jo	**Ministry of Agriculture**	**Yes**	**135**	**133**	**3**
http://www.nais-jordan.gov.jo	**Ministry of Agriculture**	**Yes***	**-**	**-**	**-**

http://www.ncare.gov.jo	**Ministry of Agriculture**	**Yes**	**10**	**10**	**-**
http://www.jordanbru.info/	**Minstry of Agriculture**	**Yes***	**-**	**-**	**-**
http://www.ncare.gov.jo/Olive Program	**Minstry of Agriculture**	**No**	**0**	**0**	**0**
http://www.acc.gov.jo	**Minstry of Agriculture**	**Yes**	**11**	**4**	**7**

*The website shows some useful information for the people in agriculture sector eg.(regulations, guidelines…ect) but doesn't show any actual dedicated service functions (electronic or nonelectronic).

In order to achieve that, a survey of these websites was conducted based on nine determinants and elements identified by Budai and Tózsa (Budai & Tózsa 2011) in their book (The Current Break through Points in Public Administration) the author chose this reference because this book talked about the Hungarian e-government and we will use the same reference in comparing between the Jordanian and the Hungarian situation in the coming articles. The nine determinants we have chosen are as follows:

Accessibility and easy availability **(A&E)**
- user friendly interface,
- simple,
- quick download,
- compatibility with Popular browsers (Explorer, Netscape and Firefox).

General basic function **(GBF)**
- searching,
- the possibility of feedback or contact (email address, phone number),
- help to use the portal,
- systematization of content according to value and depending on user's habit.

The website is flexible to change depends on the user (citizen, business, researcher, tourist, a foreigner, etc.) and this paper will check basic functions as follows:

Up to date and actualized information (**UTD&AI**)

Offering services (**OS**)

(e-) public administration services (electronic services) (**ES**)

Data security, protection of personal data (**DS**)

Possibly multilingualism (**PM**)
- choice of language (Arabic and English)

Portal functions (**PF**)
- news, events, press conferences,
- information of common interest (schedules ,cultural programs),
- basic information (local time, weather, number of visitor, etc),
- faultless operation ,quick communication.

W3C, WAI recommendations (**W3C,WAI**)

World Wide Web Consortium, Web Accessibility Initiative (http://www.w3.org/wai/): the recommendations to make the services and the websites accessible and easy to use for the mentally handicapped people or the people who live under disadvantageous conditions. In addition to the above, we must make sure of the following:

- Reading system, icon or voice control, Braille interface, changeable font size (for weakly sighted people),
- Substitution of mouse, voice controlling interface, optional waiting period during interactions (for physically handicapped people).

Moreover, the plus sign (+) shows that the website offers the function while the minus sign (-) shows it does not, the author will scan all the eight websites manually and examine the nine determinants and put (+) or (-) sings based on the Table 2 below:

Table 2. Determinants and the description (Based on Budai & Tózsa 2011)

	Determinants	**Description**
1	Accessibility and easy availability (**A&E**)	- User friendly interface (all portal functions easy to use and works) **1 mark.** - Simple (use graphics, explore easily inside the website) **1 mark.** - Quick download (downloads functions work easily) **1 mark.** - Compatibility with Popular browsers try to access with 3 main browser: Explorer, Netscape ,Firefox) **1 mark.** **The result : >=2 (+) >2 (-)**
2	General basic function (**GBF**)	- Searching. - The possibility of feedback or contact (email address, phone number). - Help to use the portal. - Systematization of content according to value and depending on user's habit. **The result : >=2(+) >2 (-)**
3	Up to date and actualized information (**UTD&AI**)	- The information and advertisements on the site are current **The result : Yes (+) NO (-)**
4	Offering services (**OS**)	- The website explains how to access to services both electronic or non-electronic. **The result : Yes (+) NO (-)**
5	Public administration services (electronic services) (**ES**)	- The site offers electronic services. **The result : Yes (+) NO (-)**
6	Portal functions (**PF**)	- News, events, press conferences. - Information of common interest (schedules, cultural programs). - Basic information (local time, weather, number of visitor, etc). - Quick communication. **The result : >=2(+) >2 (-)**
7	Possibly multilingualism (**PM**)	- The site allows you to change the language (Arabic, English). **The result : Yes (+) NO (-)**
8	Data security , protection of personal data (**DS**)	- The website mention any way of user information protection **The result : Yes (+) NO (-)**

9	W3C , WAI recommendations (**W3C,WAI**)	The author chose four major functions to verify: - Reading system, icon or voice control. - Changeable font size (for weakly sighted people). - Voice controlling interface. - Optional waiting period during interactions (for physically handicapped people). **The result : >=2(+) >2 (-)**

4. Results and discussion

After a comrehansive survey of the seven websites selected and according to the previously mentioned 9 criteria identified by the authors, the final results appear in Table 3. The authors will discuss these results for each website and there will be detailed information to these results.

Table 3. All Portals websites criteria. (Own source)

Websites / criteria	A&E	GBF	UTD	OS	ES	DS	PM	PF	W3C , WAI
https://jordan.gov.jo	+	+	+	+	+	+	+	+	-
http://moa.gov.jo	+	+	-	+	+	-	-	-	-
http://www.nais-jordan.gov.jo	+	+	+	-	-	-	+	+	-
http://www.ncare.gov.jo	+	+	+	+	-	-	+	+	-
http://www.jordanbru.info/	-	+	-	-	-	-	+	-	-
http://www.ncare.gov.jo/OliveProgram	-	-	-	-	-	-	-	-	-
http://www.acc.gov.jo	+	+	+	+	+	+	+	+	-

Abbreviations in the table: Determinants (See in Table 2)

e-Government Portal (https://jordan.gov.jo)

The e-Government Portal (Figure 6) which was launched in September 2006 in Jordan provided all the citizens, businesses, and governmental agencies with an entrance gate and an access for all the services and information offered by the government for illustration, (Alsmadi, 2011). The government has later made a decision that it is the responsibility of each ministry in the government to develop and maintain its own official website to deliver information and services (Saleh et al. 2013).

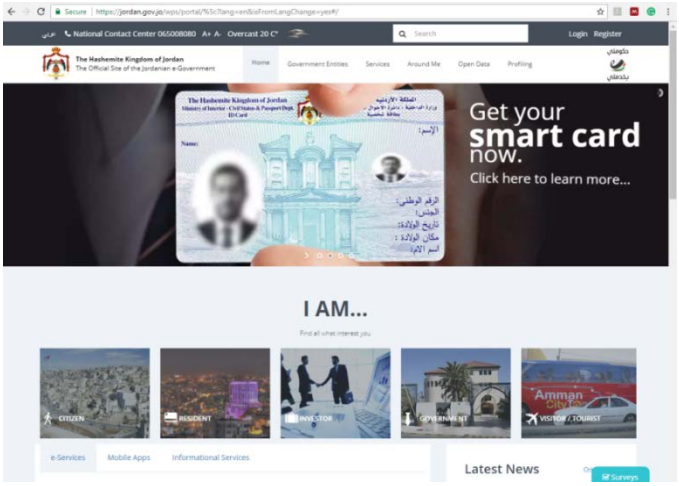


Figure 6. Jordanian e-Government Portal (www.Jordan.gov.jo) (MoICT 2016)

Furthermore, citizens and businesses nowadays have an access to about 100 e-services on the National Government Portal provided by 45 governmental agencies. These e-services included police clearance, drivers and vehicle licensing, e-Tax, a real estate registry, and national exam registration and results. There are other 50 e-services that were planned to be launched in 2016 and 70 more in 2017, according to the ICT Minister in Jordan (MoICT, 2016). Finally, The Jordanian e-Government appears on the internet through the website (https://jordan.gov.jo). And as you noticed in Table 3 this website, in general, looks good enough and has fulfilled all the criteria, but on the other hand, unfortunately, the portal is still not offering any features for the weakly sighted people or physically handicapped people. The agricultural electronic services are still of limited number and after having checked all services that this website provides just six electronic services were provided in the agriculture sector.

Ministry of Agriculture (http://moa.gov.jo)

Ministry of Agriculture was founded in 1939. The Ministry is responsible for organizing and developing the agricultural sector to achieve mainly the sustainability of production and resources without harming the environment. The Ministry is also responsible for increasing farmers' income and improve their standard of living, conservation and efficient utilization of agricultural resources and the environment and finally encourage youth people to work on the agricultural production (Ministry of Agriculture 2017b). As we can see in Table 3 above this website, in general, needs to improve the information on this website as they are out of date and it is not shown any kind of data security or any W3C, WAI recommendations, moreover it has weaknesses in the portal functions (news, events, local time weather, etc) and finally it doesn't have function to change the language (Arabic-English).

National Agricultural Information System (NAIS) (http://www.nais-jordan.gov.jo)

Reliable agricultural information plays a major role in the planning of agricultural development and formulating relevant policies. The availability of this information is important in order to enable those who are involved in the agricultural sector, whether they are institutions or individuals, to make decisions on scientific and valid bases. However, the currently available agricultural information in Jordan is still inadequate as due to many factors.

The NAIS is a Jordanian platform for information dissemination and knowledge sharing and exchange for Agricultural Research and Development (ARD) for stakeholders of the agricultural sector in Jordan. The NAIS trying to play a major link between the Ministry of Agriculture and other Agricultural stakeholders to establish an efficient and effective information system that will support agricultural development in Jordan and would strengthen and improve agricultural information generation, management, dissemination and exchange for policy-makers using web-based applications and tools, in order to:

- Support policy and decision-making in relation to national planning.
- Provide the basis for monitoring and assessing agricultural production and development.
- Support research and development, and disseminate the outputs.

- Support extension services.
- Provide an institutional memory for the MOA(NAIS 2017).

Finally after we checked the website it fulfilled 5 out of 9 cretreia the website looks simple and has a user-friendly interface but it doesn't show any kind of data security or any W3C, WAI recommendations, moreover it doesn't offer any services either electronic or non-electronic.

National Center for Agricultural Research and Extension (http://www.ncare.gov.jo)

The NCARE was established in the late 1950's when the Department of Research and Extension had been created. In 2007 NCARE was reformed to include the extension activities. The center receives primarily founds from the government and secondly from other national and international donors to support the implementation of proposed projects. The NCARE consists of a main headquarter in Amman and seven regional centers located in other Jordanian cities. It also operates (13) research stations representing different agro-ecological conditions and (13) extension field units throughout Jordan (NCARE 2017b).

In conclusion and based on results in (Table 3) this website looks good enough and has fulfilled the majority of the criteria, but unfortunately, the portal is still not offering any features for the visitors with specific needs (weakly sighted people or physically handicapped) and the site offers just 10 offline services (the user has to appear physically), it does not offer any electronic services.

Bee Research Department (http://www.jordanbru.info/)

Bee Research Department was established in 2001, the Department is the first of its kind in Jordan and the Middle East, it includes a group of modern and unique laboratories such as pollen grains studies laboratory and honey bee diseases identification using biotechnology techniques laboratory in addition to the honey bee hereditary bank. The bee research department was documented as a global source for the local honey bees (NCARE 2017a). It appears on the internet. Unfortunately, this website fulfilled just 2 out 9 criteria (Table 3) and when we go through this website we can see it looks unuseful with a complex interface and doesn't show any kind of data security or any W3C, WAI recommendations, moreover it doesn't offer any services even electronic or non-electronic or changing language function, so in conclusion, it needs to restructure.

Olive Research and Extension Program in Jordan (http://www.ncare.gov.jo/OliveProgram)

In view of the importance of the olive tree in the Jordanian agricultural sector in general as it covers 36% of the total cultivated area in Jordan (86100 hectares) according to statistics of the Ministry of Agriculture for 2014. So, this department has been established in 2002 to improve the quantity and quality of olive production, build an information base for the olive sector and transfer the results of olive research to farmers by all possible means (NCARE 2017c). However, the appearance of this department on the internet still so shy.and this website looks the worse between all websites and as you can see in (Table 3) it didn't fulfill any of criteria mentioned before, it looks unuseful with a complex interface, out of date information and doesn't show any kind of data security or any W3C, WAI recommendations, moreover it doesn't offer any services neither electronic or non-electronic or changing language function, so in summary, it needs to restructure.

Agricultural Credit Corporation (http://www.acc.gov.jo)

The idea of agricultural lending goes back to the Othman empire before the existence of Jordan. However, after the establishment of the state of Jordan, this institution went through various changes until 1960 when the Jordanian government decided to merge all institutions acting in this field into one institution called the Agricultural Credit Corporation (Ministry of Agriculture 2017a). This website fulfilled 8 out of 9 criteria and it seems one of the best websites we have checked with a simple and friendly interface, up to date information and most of the criteria we have, moreover, it offers 4 electronic services. The website is still not offering any features for the weakly sighted people or physically handicapped people (reading system, icon or voice control, voice controlling interface, optional waiting period during interactions).

Discussion

After conducting this survey and analysis of all the websites dealing with the agriculture sector, the results were as following:

- Most websites showed easy access and use interfaces, simple and compatibility with popular browsers except two websites.
- In general basic function, all websites showed flexibility to change depending on the user and basic functions like searching, the possibility of feedback or contact (email address, phone number) and help to use the portal except for only one site.
- Up to date and actualized information, only 50% of these websites have continuously updated the information on their sites.
- Offering services (offline and online): there were 4 websites out of 7 that provided services across the site.
- Only 3 websites have offered these services electronically.
- Data security and protection of personal data, only two websites have been established and clarified to protect users' information and data.
- Most of the websites offers the change of language (Arabic and English).
- Portal functions: 4 websites of 7 showed such features as
    - News, events, press conferences.
    - Information of common interest (schedules, cultural programs).
    - Basic information (local time, weather, number of visitors, etc).
- Finally, unfortunately, no site has shown services that are in line with W3C, WAI recommendations such as reading system, icon or voice control, Braille interface, changeable font size (For weakly sighted people) or substitution of mouse, voice controlling interface, optional waiting period during interactions (for physically handicapped people).

5. Conclusion

In conclusion, many studies have shown the importance of the use of technology in agriculture and the importance of the use of the internet in communication and provision of services in the agricultural sector. Official portals are the facade of government institutions, such as mirror, reflects the communication between these institutions and users. In this paper we made a survey for the institutions on the internet that provide agricultural services to know the current status of these websites and readiness to provide electronic services for business and workers in the agricultural sector. These websites still under the required level and need comprehensive review in future. For example, it has been shown that most of them do not provide electronic services and even the number of electronic services provided through these websites are still limited (no more than 10 services). These portals did not show high-security tools for protecting user information and personal data. In addition to it, these websites do not have any features for people with special needs that allow them to use these websites.

Finally this paper is the first phase of the research to study the current situation of the electronic services in the agricultural sector in Jordan. In the second stage of the research, there will be a collection phase to find the successful experiences from the European Union and an attempt to improve the current situation of the electronic services in Jordan of this sector. Hopefully these selected services could lead to the development in Jordan’s agriculture. Further research will go deeper into these electronic services and know how to increase them and explore the level of acceptance for these services in Jordan.

References

Alsmadi, I 2011, 'Security Challenges For Expanding E-governments’ Services', International Journal of Advanced Science and Technology, 37.

Botos, Sz, Herdon, M & Várallyai, L 2015, 'Readiness for Future Internet Services in Rural Areas', Procedia Economics and Finance, 19, pp. 383–390. https://doi.org/10.1016/S2212-5671(15)00039-8

Budai, BB & Tózsa, I 2011, The Current Break through Points in Public Administration First edit., Covrinus University Budapest.

Córdova, FM, Durán, C & Galindo, R 2014, 'Comparative analysis of ICT in public–private systems: The OHIM

case in the European Union and the Internal Revenue System in Chile', Procedia - Procedia Computer Science, 31(31), pp. 95–104. Available at: www.sciencedirect.com [Accessed March 24, 2018].

Deichmann, U, Goyal, A & Mishra, D 2016, 'Will digital technologies transform agriculture in developing countries?', Agricultural Economics, 47(S1), pp. 21–33. Available at: http://documents.worldbank.org/curated/en/481581468194054206/pdf/WPS7669.pdf [Accessed March 25, 2018].

Department of Statistics, J 2017, Department of Statistics., Available at: http://dosweb.dos.gov.jo/ [Accessed December 31, 2017].

Ebrahim, Z & Irani, Z 2005, 'E-government adoption: architecture and barriers', Business Process Management Journal, 11(5), pp. 589–611.

FAO 2017, Information And Communication Technologies For Agriculture And Rural Development, Available at: http://www.fao.org/3/a-i6733e.pdf [Accessed March 26, 2018].

FAO and International Telecommunication Union 2016, E-Agriculture Strategy Guide Piloted in Asia-Pacific countries,

Gatautis, R, Medziausiene, A, Tarute, A & EV 2015, Towards ICT Impact Framework: Private and Public Sectors Perspective. Available at: http://www.joebm.com/papers/229-A10016.pdf [Accessed March 24, 2018].

Gelb, E & Parker, C 2005, 'Is ICT Adoption for Agriculture Still an Important Issue ?' In ICT in Agriculture: Perspectives of Technological Innovation. pp. 1–8. Available at: http://departments.agri.huji.ac.il/economics/gelb-gelb-parker.pdf [Accessed March 25, 2018].

Gil-García, JR & Pardo, TA 2005, 'E-government success factors: Mapping practical tools to theoretical foundations', Government Information Quarterly, 22(2), pp. 187–216.

Jordan Times 2017, ICT sector contributes. Jordan Times.

Karetsos, S 2014, 'Developing a smartphone app for m-government in agriculture', Journal of Agricultural Informatic, 5(1), pp. 1–8. Available at: http://real.mtak.hu/23955/1/129_615_1_PB_u.pdf.

Lengyel, P. et al., 2015, 'Social network analysis of international scientific collaboration on psychiatry research' International Journal of Mental Health Systems, 9(1).

Lin, F, Fofanah, SS & Liang, D 2011, Assessing citizen adoption of e-Government initiatives in Gambia: A validation of the technology acceptance model in information systems success. Government Information Quarterly, 28(2), pp. 271–279.

Ministry of Agriculture 2017a, Agricultural Credit Corporation. Available at: http://www.acc.gov.jo [Accessed December 31, 2017].

Ministry of Agriculture 2017b, Ministry of Agriculture. Available at: http://moa.gov.jo/ar-jo/home.aspx [Accessed December 31, 2018].

MoICT 2016, E-Government Program. Available at: http://www.moict.gov.jo/Pages/E-Government-Program-Objectives.aspx [Accessed December 15, 2016].

NAIS 2017, National Agricultural Information System. Available at: http://www.nais-jordan.gov.jo [Accessed December 31, 2018].

NCARE 2017a, Bee Research department. Available at: http://www.jordanbru.info/ [Accessed December 30, 2017].

NCARE 2017b, National Center for Agricultural Research and Extension. Available at: http://www.ncare.gov.jo [Accessed December 30, 2017].

NCARE 2017c, Olive research and extension program in Jordan. Available at: http://www.ncare.gov.jo/OliveProgram%0D%0A [Accessed December 30, 2017].

Ntaliani, M. et al., 2010, 'Agricultural e-government services: An implementation framework and case study', Computers and Electronics in Agriculture, 70(2), pp. 337–347. Available at: http://www.sciencedirect.com/science/article/pii/S0168169909001884 [Accessed December 31, 2017].

Rahim, AA & Athmay, A 2013, Citizens ' Perceptions towards e-Governance : Field Study , 7(9), pp. 631–639.

Saleh, ZI, Obei, RA & Khamayseh, Y 2013, 'A Framework for an E-government Based on Service Oriented Architecture for Jordan', I.J. In formation Engineering and Electronic Business, 3(3), pp. 1–10.

Saravanan, R 2010, ICTs for Agricultural Extension: Global Experiments, Innovations and Experiences, New india publishing.

Seri, DP & Zanfei, A 2016, 'The role of ICT, skills and organizational change in public sector performance', Argomenti, pp. 5–30.

Statista 2016, Mobile phone penetration worldwide 2013-2019 | Statistic. Mobile phone user penetration as percentage of the population worldwide from 2013 to 2019. Available at: https://www.statista.com/statistics/470018/mobile-phone-user-penetration-worldwide/ [Accessed March 25, 2018].

Szilágyi, R & Szilágyi, E 2009, 'E-government readiness - where are we going ? A short overview of E-government in Hungary', 4th Aspects and Visions of Applied Economics and Informatics, (March).

United Nations 2002, Benchmarking E-government: A Global Perspective Assessing the Progress of the UN Member States,

We Are Social 2017, Digital in 2017: Global Overview - We Are Social. We Are Social. Available at: https://wearesocial.com/special-reports/digital-in-2017-global-overview [Accessed March 26, 2018].

Hungarian Association of Agricultural Informatics
European Federation for Information Technology in Agriculture, Food and the Environment

Journal of Agricultural Informatics. Vol. 9, No. 2
journal.magisz.org

Application of the multiple objective programming in the optimization of production structure of an agricultural holding

Lajos Nagy[1], László Pusztai[2], Margit Csipkés[3]

I N F O
Received 2 Mar 2018
Accepted 7 May 2018
Available on-line 18 Jun 2018
Responsible Editor: M. Herdon

Keywords:
linear programminf, MOLP, agricultural, sequential programming.

A B S T R A C T

At an agricultural plant during the preparation of the annual plan, besides taking resources as well as market opportunities into account, we wish to accomplish a production structure which can provide maximum income for the company. As an effect of climate change extreme weather conditions can be experienced more frequently than before, which conditions are tolerated in different ways by arable cultures with variable terrain features. When there is an agreement in the final production structure, decision makers take even risk factors into account in their decision making. Endeavour for reducing burden on the environment is playing a more and more important role in decision making. These endeavors are often of opposite directions but they can be coordinated as well as compromises can be found by the application of multiple objective programming. In our article we aim at introducing opportunities for the application of this method.

1. Introduction

Optimization is generally carried out by aiming at only one goal in programming models (Winston 1997). In economic models the most common goals are either those of maximizing income or minimizing costs. In some cases, however, the decision maker needs to set up more than one different goals simultaneously. For instance, a manufacturing company would like to focus on utilizing their available resources in the most efficient ways, which requires achieving several and counteractive goals at the same time. For instance, when a company would like to reach, several, opposing goals at a time. Such counteractive goals are those of maximizing income and minimizing cost parallelly. Furthermore, a small- and medium sized company can have social political goals as well in order to provide a relatively high employment ratio integrated with the two previous goals. In order to protect the environment, the reduction of pollutant emission, the increase in the level of customer service, - and the list could be further expanded -, are all important goals in a company's life, but these goals can be counteractive in both simple or complicated ways. The harmonization of these different goals is not always an easy task because the quantificated values of these goals can be one of the sources for a company's competitiveness (Ragsdale 2007).

Multiple objective programming is widely applied in the field of economics and finance, in the optimization of production processes as well as production structure optimization and in many other fields of life. The applied methods are also quite diverse as a wide range of operations research methods are applied in solving different problems.

Multiple objective programming was applied to solve site location problems (Scheiderjans et al. 1982) or snow removal work in Montreal was modelled by Cambell & Langevin (1995). Chih and

[1] Lajos Nagy
University of Debrecen
nagy.lajos@econ.unideb.hu
[2] László Pusztai
University of Debrecen
pusztai.laszlo@inf.unideb.hu
[3] Margit Csipkés
University of Debrecen
csipkes.margit@econ.unideb.hu

Ching (1996) applied multiple objective programming to create water quality management. Agricultural applications are also quite widespread: Tóth & Szenteleki published their work on compromise programming in 1983. However, we can read about agricultural research in the field of cattle breeding (Mendez et al. 1988), in regional agricultural forecast (Zekri & Romero 1992), as well as in vegetable production (Berber et al. 1991). Risk programming models combine the elements of linear programming or quadratic programming with those of game theory to resolve conflicts between profitability and risk as well as to select the compromise most suitable for the decision maker (Hazell 1971; Hazell & Norton 1986; Hardaker et al. 1997; Berbel 1993).

In our article, we look for answers to whether a plant-growing enterprise can successfully use multi-purpose programming at decision-making. We review basic model types, present a real-life practical application, and provide suggestions on how to implement and choose models.

2. Multiple Objective Linear Programming models (MOLP)

Numerous papers have been published on several versions and algorithms of multi objective programming. Article written in 2015 by Colapito et al. provides a detailed overview and summary about the emergence and evolution of models as well as their applicability in different professional fields.

In this chapter we are going to introduce the basic types of the above mentioned models, and their role in analyses. We are going to deal with goal programming and MOLP model in detail, because it was these methods that we applied to determine the optimal production structure with regard to multiple targets at an agricultural plant.

2.1. General determination and justification of the model

$$x_j \geq 0 \quad j = 1,2,3,..n$$

$$\sum_j a_{ij} x_j \leq b_i \qquad x_j : \text{var}ied$$

$$\sum_j a_{ij} x_j \geq b_i \qquad a_{ij} : specific\ need$$

$$\sum_j a_{ij} x_j = b_i \qquad b_i : confine$$

K K K K K K K

$$\sum_j c'_{kj} x_j = optimum$$

$$\sum_j c'_{2j} x_j = optimum \qquad c'_{kj} : objective\ coefficient\ value$$

In the general model we set the constraints as well as the goals preferred by the decision maker. In case of contradictory goals (e.g. cost reduction and maximum income achievement), the enhancement of one goal will entail the enhancement of the other, which may not be desirable. For example, in particular areas, such as the different branches of plant cultivation as well as in the selection of farming intensity, apart from profitability aspects, expenses play an important role as well, especially in case of less capital-intensive farming companies. Therefore, regarding goals, producers have to make compromises taking their circumstances into account.

Compromises are a necessary part of the decision making process, however, in case of conventional planning it is not easy to find a solution, which results in making only minimal concessions in the goals (Figure 1).

Point A indicates the possible minimal cost while achieving goals. The possible maximum income is indicated by point B. The choice of point C is not logical for the decision maker, because if we take the first goal into account, there is a solution, where regarding target costs, the first target value (point A) is lower than the second target value (point C). In addition, there is a solution, where regarding

target income, the second target value (point B) is higher, than the first target value (A). As a consequence, solutions A and B always dominate over C.

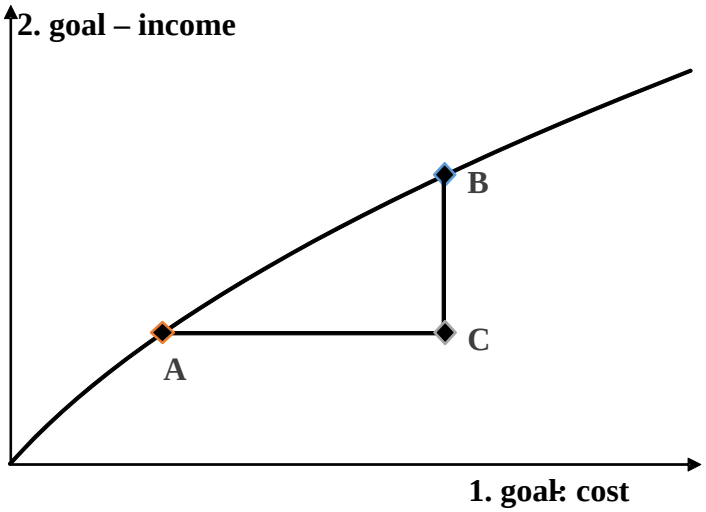


Figure 1. Compromise decisions – dominant and non-dominant versions (Source: Own editing according to Ragsdale, 2007)

Multi objective programming can guarantee non-dominant solutions and it enables to make decisions at system level. Every solution in the trade-off curve section between A and B is Pareto-optimal.

2.2. Modelling opportunities in case of multiple objectives

2.2.1. Alternative programs

Alternative programming can be applied when a compromise solution is calculated between two, one directional goals (Chiandussi 2012). One of our goals can be the achievement of maximal gross margin, the other can be the achievement of maximal net income. The method is quite simple. First of all, the linear programming model is solved on the basis of the first, then of the second objective. In case different optimal values are calculated, the programs can be combined with distribution ratios (Table 1).

Table 1. Compromise solution determination with the use of alternative linear programming

	Optimum 1. objective	**Optimum 2. objective**	**Compromise**
Variables	x_{1opt1}	x_{1opt2}	$k_1x_{1opt1}+k_2x_{1opt2}$
	x_{2opt1}	x_{2opt2}	$k_1x_{2opt1}+k_2x_{2opt2}$
	⋮	⋮	
	x_{mopt1}	x_{mopt2}	$k_1x_{mopt1}+k_2x_{mopt2}$
$z_1$max	z_{1opt1}	z_{1opt2}	$k_1z_{1opt1}+k_2z_{1opt2}$
$z_2$max	z_{2opt1}	z_{2opt2}	$k_1z_{2opt1}+k_2z_{2opt2}$
Distribution ratios	k_1	k_2	$k_1+k_2=1$

2.2.2. Sequential programming

In case of sequential programming we rank goals according to their importance. Optimization is started with the most important goal: $f_1(x) = \sum_j c'_{1j} x_j$, with an L_1 solution set. Subsequently, optimization with the remaining goals is carried out in accordance with the obtained importance rank to get $L_2, L_3, \ldots L_m$ solution sets. If there is a common L, which is true for $L \subset L_{m-1} \subset \ldots \subset L_2 \subset L_1$, it means that all the objective functions have their own optimum value within the L solution set, otherwise all the goals cannot be optimized simultaneously.

The method is quite simple. Although its efficiency is questionable, it definitely has at least two advantages. On the one hand, goals with similar optimum values can be identified, which makes it possible to reduce the number of objective functions for further analysis. On the other hand, marginal values of objective functions (e.g. resources, market conditions, etc.) can be revealed and it plays an important role in further examinations (Benayoun et al. 1970).

2.2.3. Constraints programming

The most important goal will be the objective of this model, all the other goals are considered as constraints, p_i constant is on the right side of the equation. These constants are between the predetermined minimal (m_i) and maximal (M_i) goal values ($m_i \le p_i \le M_i$). Objective function values of secondary goals acquired by sequential programming can provide a useful point of reference for the determination of p_i.

The model:

$$
\begin{array}{l}
x_j \ge 0 \quad j = 1, 2, 3 \ldots n \\
\sum_j a_{ij} x_j \le b_i \\
\sum_j a_{ij} x_j \ge b_i \\
\sum_j a_{ij} x_j = b_i \\
\left.\begin{array}{l}
\sum_j c'_{1j} x_j = p_1 \\
\sum_j c'_{2j} x_j = p_2 \\
\qquad \text{M} \\
\sum_j c'_{kj} x_j = p_k
\end{array}\right\} \quad \text{by-end} \\
\hline
\sum_j c'^{\,\text{főcél}}_{j} x_j \to \textit{extrém}
\end{array}
$$

In case of secondary goals, the relations are not necessarily equalities. There can be either $\le$ (minimal goals) or $\ge$ (maximum goals). With this method our main goals within the given restrictions will be definitely accomplished, in case of all the other goals - with regard to deviations from them - we can carry out further research through sensitivity analysis (Marler & Arora 2004).

2.2.4. Goal programming

Instead of objective functions, we insert equations with predetermined goal values into the constraints. The objective function minimalizes the sum of negative or positive deviations from the targeted values. Constraints regarding goals:

$$d_1^- \geq 0, \quad d_1^+ \geq 0$$
$$\sum_j c'_{1j} x_j + d_1^- - d_1^+ = t_1$$
$$\sum_j c'_{ij} x_j + d_i^- - d_i^+ = t_i$$
$$\text{M}$$
$$\sum_j c'_{kj} x_j + d_k^- - d_k^+ = t_k$$

The objective function:

$$\sum_i \boldsymbol{d_i^- + d_i^+ \rightarrow MINIMUM}$$

Function no. 4. can be used when the measure unit of goals are identical and when there are not disturbing differences between their quantities. On the contrary, it is practical to calculate with the relative deviation from the targeted goal that can be expressed in percentages:

$$\sum_i \frac{1}{\boldsymbol{t_i}} (\boldsymbol{d_i^- + d_i^+}) \rightarrow \boldsymbol{MINIMUM}$$

The question arises as to how individual goals can be ordered on the basis of their importance, because there can be goals, whose deviations from the target goal lead to bigger differences. In this case weights can be added to deviation variables:

$$\sum_i (w_i^- d_i^- + w_i^+ d_i^+) \rightarrow MINIMUM$$

or

$$\sum_i \frac{1}{t_i} (w_i^- d_i^- + w_i^+ d_i^+) \rightarrow MINIMUM$$

By applying goal programming, the fine adjustment of different goals can be made possible. By using weights, one or more goals can be highlighted, and the decision maker will have the opportunity to find the most appropriate compromise solution (Ragsdale, 2007).

2.2.5. MOLP – Multiple objective linear programming

With the use of goal programming we searched for such a compromise solutions, where the total deviation from goals were minimal. MOLP method provides us another solution. In this case, the goal is to find the minimal value of the deviation from individual goals. First of all, deviations from individual goals must be calculated:

$$\frac{\sum_j \boldsymbol{c'_{ij} x_j - t_i}}{\boldsymbol{t_i}}$$

This can be weighted in accordance with their importance, like in the case of goal programming:

$$\boldsymbol{w_i} \left(\frac{\sum_j \boldsymbol{c'_{ij} x_j - t_i}}{\boldsymbol{t_i}} \right)$$

$\boldsymbol{Q}$ minimax variable was implemented, which is constraints as well. Thus, the objective function of the model:

$$\boldsymbol{Q \rightarrow MINIMUM}$$

which has the following constraint:

$$\boldsymbol{w_i} \left(\frac{\sum_j \boldsymbol{c'_{ij} x_j - t_i}}{\boldsymbol{t_i}} \right) \leq \boldsymbol{Q}$$

According and optimal solution can be calculated, where the maximum deviation from individual goals is the lowest. With this method fault can be avoided that the total deviation is minimal, but there are poorly-performing objectives, which are common in goal programming (Winston 1997). The question can arise: whether the usage of goal programming or MOLP is more expedient? There is no exact answer for this question, but it is a fact, that results calculated with goal programming are belongs to an extremal point, while the results of MOLP is not in every case. This can usually result in that the deviation from individual goals is lower when using MOLP, so this can mean a good estimation (Figure 2).

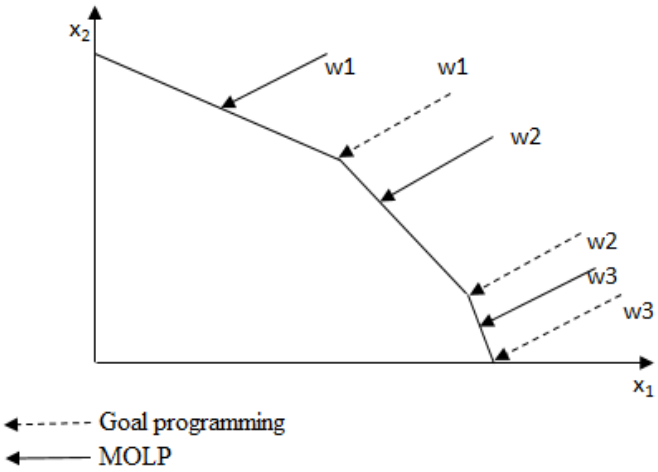


Figure 2. Solutions acquired by goal programming and MOLP with the use of different weights

3. Production ratio optimization at a crop producer in Hajdúság based on several objectives

The company in target utilizes a 2000-acre farming area: they produce winter wheat, corn, oilseeds (like sunflower, rape) and in irrigated areas they grow peas as well. Thus the variables of the model are: corn (x_1), sunflower (x_2), winter wheat (x_3), rape (x_4) and peas (x_5). The production ratio is optimized considering the following objectives: Revenue, Income, Sectoral result per 100 HUF production cost and Production cost. Objective factors are shown in Table 2.

Crop rotation conditions were considered as constraints in the model. Corn can be produced in the same area after every second year, while sunflower, rape and peas can be grown in the same soil only after every fifth year. Wheat can occupy maximum 60% of the total production area. The capacity of irrigation is 250 hectares. Regarding machines, professional work and unskilled work, the specific resource requirements were determined based on decade-specific technologies, while the amount of resources accessible in certain periods was given in hours.

Table 2. Sectoral indicators related to targeted objectives

Objectives	**Corn**	**Sunflower**	**Winter wheat**	**Rape**	**Peas**
Revenue (HUF/hectare)	436 800	230 000	266 900	378 000	684 000
Income (HUF/hectare)	152 750	103 632	82 096	136 984	266 000
(Income per production cost) x 100	45.73	58.76	34.96	47.07	56.84
Production cost (HUF/hectare)	334 050	176 368	234 804	291 016	468 000

The following model variants were calculated and evaluated:

- Sequential programming

The models were calculated for each objective separately. There were two reasons to apply sequential programming. On the one hand, the possible extreme values and related optimal values for each objective were in the focus of the research. On the other hand, filtering common solution sets were also aimed at.

- Goal programming

Goal programming model was calculated with both absolute and relative weights. The model's objectives were those individual objectives which was calculated by sequential programming. 5 variants were calculated for each model. The difference between variant were in the importance weights of objectives. In the first variant, all objectives has the same weight. In the cases of the following variants production cost objective was get higher weights, which means that in the case of fifth variant, the weight become 5.

- MOLP model

During the calculation of deviations from the goals, the extreme values acquired by sequential programming were used, and we assigned weight in accordance with the focus on the comparability with the results of goal programming.

2.3. Results of sequential programming

During sequential programming 4 model variants were created. The model variants differed from one another with respect to the highlighted objectives (Table 3).

Table 3. Model variants

Name of the variant	Highlighted objectives
Sz1	Maximum revenue
Sz2	Maximum income
Sz3	Maximum (income per production cost) x 100
Sz4	Minimum production cost

In the cases of Sz2 and Sz3 variants, programs and objectives are the same, so the second (income) and the third ((Income per production cost) x 100) objectives can be optimized at the same time (Table 4). Therefore, the third objective (income per 100 HUF production cost) is excluded from goal programming and MOLP models.

Table 4. Results calculated by sequential programming per variants

Name		Variants			
		Sz1	**Sz2**	**Sz3**	**Sz4**
Crop ratio	**Corn (ha x 100)**	5.77	5.45	5.45	5.77
	Sunflower (ha x 100)	2.15	4.00	4.00	4.00
	Winter wheat (ha x 100)	6.60	5.12	5.12	5.61
	Rape (ha x 100)	4.00	4.00	4.00	4.00
	Peas (ha x 100)	1.48	1.43	1.43	0.62
Objectives	**Revenue (Thousand HUF)**	**730 008**	715 801	715 801	687 442
	Income (Thousand HUF)	258 742	**259 602**	259 602	246 961
	(Income per production cost) x 100	45.3	46.7	**46.7**	45.7
	Production cost (Thousand HUF)	571 266	556 198	556 198	**540 480**

Rape achieves the production area constraint level in every variant, while corn achieves the maximum production area in the second, third and fourth variants. The area of the other plants remained under the constraint level in every objective. The production area of the corn (557 hectares) are equal for both the maximum revenue and minimum production cost, and it has a somewhat less significant role if the maximum of the sectoral result is calculated (545 hectares). Winter wheat has the

largest production area (660 hectares) in the case of maximum revenue objective, while this result is 100 hectares less in the case of production cost objective, however, winter wheat takes up about 150-acre smaller area in the calculation of the maximum of sectoral result. In the case of the first and second objectives it is advisable to grow peas in almost the same size of areas (143 hectares), however, its cost claim reduces its competitiveness (62 hectares) (Table 4).

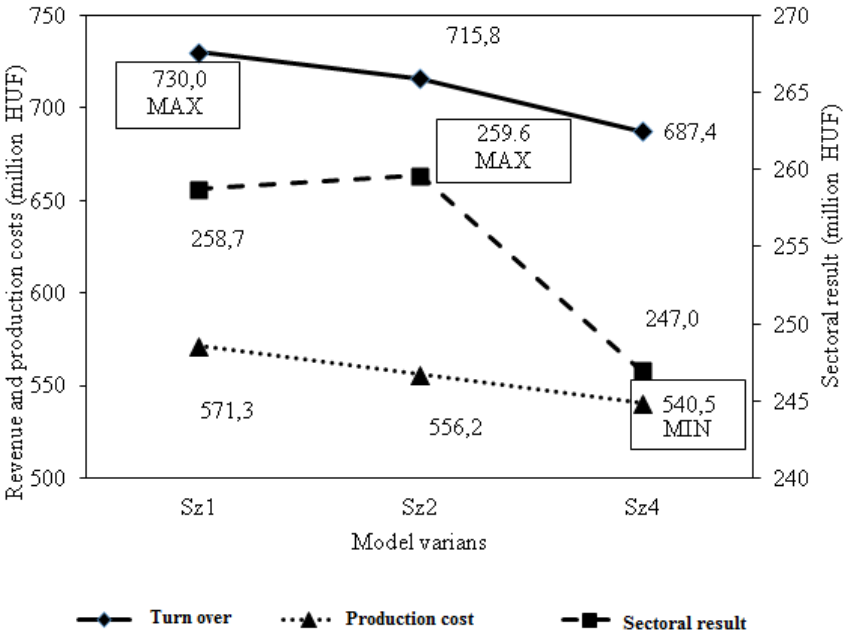


Figure 3. Objectives values calculated by sequential programming in the cases of different objectives

There is no significant difference between the sectoral result achieved at maximal revenue and the maximum sectoral result, however, the revenue related to the maximum sectoral result is approximately 15 million HUF less than the achievable maximum revenue. In the case of the achievable minimum cost, both the revenue and income decreased as expected (with 42.6 and 12.6 million HUF). Revenue loss in this case is 5.8%, result decrease is 4.8% compared to the maximum possible values

In the following we examined, what kind of opportunities are exist to search for compromise with the use of goal programming and MOLP. We considers available extreme values of goals as objectives.

2.4. Evaluation of results acquired by goal programming and MOLP method

During goal programming, the models were calculated by using both relative and absolute weights. These models provided us the same results, therefore, in the comparison section only the relative weight results will be presented.

During calculation, for all objectives, deviations from the goal was taken into account with the same weights, then production cost was highlighted by penalty weights. Weights were increased from 1 to 5, and production cost as a goal is get bigger weights. These calculations were repeated in both goal programming and MOLP model. As a first step the results were compared, then the effects of increasing weight's of production cost were evaluated.

According to Table 5 there is no significant difference between the results of goal programming and those of the MOLP method. The result calculated by goal programming is the same as in the sequential model, where the maximum sectoral result was calculated. The total absolute deviation from the goals is 29.9 million HUF, the revenue loss is 14.2 million HUF, the cost increase is 15.7 million HUF compared to the individual optimal solutions. Seemingly, the performances of MOLP are worse, because the total absolute deviation from the goals is 3.3 million HUF more, and in the cases of revenue goals, the performance of the sector result is worse than that of the goal programming model.

Table 5. Values and deviations calculated by goal programming and MOLP model

Unit: million HUF

Goal	Goal programming	MOLP	Target value
	Result		
Revenue	715.8	711.9	730.0
Income	259.6	258.0	259.6
Production cost	556.2	553.9	540.5
	Deviation from the goal		
Revenue	-14.2	-18.1	
Income	0.0	-1.6	
Production cost	15.7	13.4	
Total deviation	29.9	33.2	

It is essential that the goals should not be in the same scale in absolute value, which means that using relative deviations could provide another picture. Based on the relative deviations from the targets, similar observations can be made as in the case of absolute deviations, that is, the higher difference there is between the revenue and the result of the MOLP model, the lower production cost there will be. There is a remarkable difference in Table 7. The total relative deviation is higher in the case of the MOLP model (Table 7, w1 column). The highest relative deviation can be identified in goal programming. This result was expected because in the objective of the MOLP model, the biggest relative deviation was minimized (Figure 4).

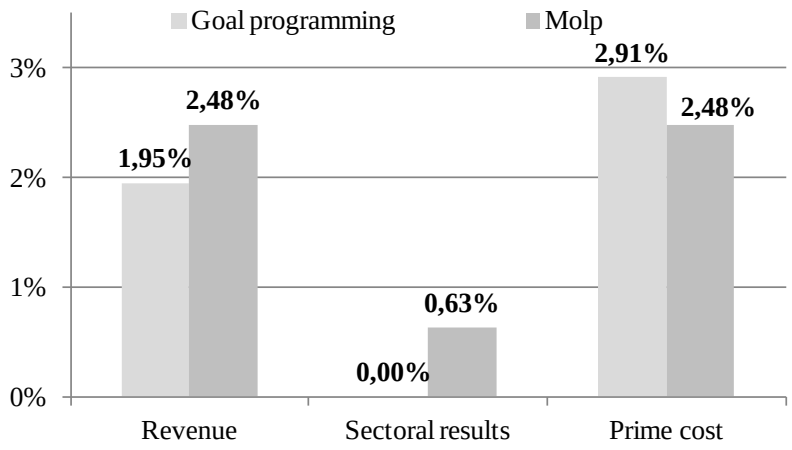


Figure 4. Relative deviations from the goal in the cases of goal programming and MOLP method

Table 6. Production structure

Unit: hectare

	Goal programming	MOLP
Corn	545	522
Sunflower	400	400
Winter Wheat	512	535
Rape	400	400
Peas	143	143

Considering production structure, the production areas of corn and winter wheat are not constant, while the production areas of other plants are of the same size regarding the results of both models (Table 6).

The next step was the assignment of penalty weights to production costs. This means that production cost as a goal is becoming more and more important compared to other goals. The role of penalty weights is easy to understand, because in the case of goal programming, by increasing the weight from 1 to 2, w=2 being twice as much as in w=1 case, assuming the same production structure. As we search for the minimum of the total relative deviation, the optimal program will only be changed if the optimum belongs to another extremal point (Figure 2.) In the cases of goal programming, there is no change in w=1 and w=3 weights, only the linear increase of the relative deviation of production cost can be seen (2.9% ~~5~~.8%~~8~~.7%). The deviation of the revenue and income did not change (1.9% and 0.0%). We experienced some changes with w=5 and w=6 weights. In this case, further increases in the weights would induce such a big change in the objective function that the optimal solution would belong to another extremal point.

In the case of the MOLP model it is revealed that with the increase of penalty weights, the total relative deviations will differ from. In the production cost, a slow decrease, while in the cases of the other two goals continuous increase can be detected (Table 7).

Table 7. Relative deviations from targets in case of 1-5 production cost penalty weights

Goals	**Goal programming**					
	w1	**w2**	**w3**	**w4**	**w5**	**MAX**
Revenue	1.9%	1.9%	1.9%	5.8%	5.8%	5.8%
Income	0.0%	0.0%	0.0%	4.9%	4.9%	4.9%
Production cost	2.9%	5.8%	8.7%	0.0%	0.0%	8.7%
Sum	4.9%	7.8%	10.7%	10.7%	10.7%	
	MOLP					
	w1	**w2**	**w3**	**w4**	**w5**	**MAX**
Revenue	2.5%	3.5%	4.0%	4.3%	4.6%	4.6%
Income	0.6%	1.8%	2.5%	3.0%	3.3%	3.3%
Production cost	2.5%	1.7%	1.3%	1.1%	0.9%	2.5%
Sum	5.6%	7.0%	7.9%	8.4%	8.8%	

Comparing the results of the two models in detail, the initial disadvantage of MOLP disappears with the increase of penalty weight, the total relative deviation is lower in the case of w=2, then goal programming. If we take the deviations from goals, MOLP seems more balanced.

Table 8. Change in production structure, in the cases of production cost penalty costs

Unit: hectare

	Goal programming				
	w1	**w2**	**w3**	**w4**	**w5**
Corn	545	545	545	577	577
Sunflower	400	400	400	400	400
Winter wheat	512	512	512	561	561
Rape	400	400	400	400	400
Peas	143	143	143	62	62
	MOLP				
	w1	**w2**	**w3**	**w4**	**w5**
Corn	522	507	523	533	540
Sunflower	400	400	400	400	400

Winter wheat	535	561	561	561	561
Rape	400	400	400	400	400
Peas	143	132	116	106	99

Optimal programs presented in table 8 verify the above statements. The goal programming basic model's result (where the total weight is 1) equals with that solution, where sectoral result maximum was calculated. In the cases of those variants, where production cost penalty weights were w=4 and w=5 the optimal program is the same as sequential model, where the objective was the production cost. Changing the weight results in the selection of these two models.

In the cases of MOLP models, with the increase of the importance in production cost, tendencies in the production structure is similar like in the goal programming model. The production area of corn and winter wheat increases, the production area of peas decreases, while the areas of rape and sunflower are on the upper constraint in every variant. Optimal programs do not connect to extremal points, but all of them are on the trade-off curve.

3. Conclusion

In practice managers have to make decisions considering several objectives. One is more important than the other, nevertheless, none of these objectives should be excluded from the final decision making.

In the article some application opportunities and the importance of Multiple Objective Linear Programming methods are examined. The applicability of goal programming and that of MOLP were compared through an example of an agricultural enterprise.

It is suggested as a first step that the opportunities per goals be analyzed with the use of sequential programming. The calculated results can only provide information on single objectives, however, they can be applied during further examination. By using sequential programming, simultaneously optimizable objectives can be filtered, which can also help to make further examinations easier.

In the next step the application of both goal programming and MOLP models can be considered. In both models, we can represent the importance of high-priority goals with their assigned penalty weights. During our research, we have found that the goal programming model is less sensitive to the change in penalty weights than the MOLP model, so if the relative importance of our goals is different (for example, the priority of cost cutting is outstanding), it is more appropriate to use MOLP.

References

Benayoun R, de Montgolfier J, Tergny J & Laritchev O 1970, 'Linear programming with multiple objective functions: Step method (stem)' Mathematical Programming, Volume 1, Issue 1, pp. 366–375.

Berbel J 1993, 'Risk programming in agricultural systems: A multiple criteria analysis' Agricultural Systems, Volume 41, Issue 3, pp. 275-288. https://doi.org/10.1016/0308-521X(93)90004-L

Berbel J, Gallego J & Sagues H 1991, 'Marketing goals vs. business profitability: An interactive multiple criteria decision-making approach' Agribusiness,vol.7,issue 6, pp. 537-549. https://doi.org/10.1002/1520-6297(199111)7:6<537::AID-AGR2720070604>3.0.CO;2-D

Cambell J & Langevin A 1995, 'The Snow Disposal Assigment Problem' The Journal of the Operational Research Society. Vol 46. No.8. pp. 919-929. https://doi.org/10.1057/jors.1995.131

Chih-Sheng Lee & Ching-Gung Wen 1996, 'Application of Multiobjective Programming to Water Quality Management in a River Basin' Journal of Environmental Management. Vol. 47. Issue 1. pp. 11-26. https://doi.org/10.1006/jema.1996.0032

Chiandussi, G. Codegone M, Ferrero S, Varesio FE, 2012, 'Comparison of multi-objective optimization methodologies for engineering applications' Computers & Mathematics with Applications, Volume 63, Issue 5, Pages 912-942. https://doi.org/10.1016/j.camwa.2011.11.057

Colapinto C, Jayaraman R & Marsiglio S 2015, 'Multi-criteria decision analysis with goal programming in engineering, management and social sciences: a state-of-the art review' Annals of Operations Research, Online First 1-34 (http://link.springer.com/article/10.1007%2Fs10479-015-1829-1)

Hazell PBR 1971, 'A Linear Alternative to Quadratic and Semivariance Programming for Farm Planning Under Uncertainty' American Journal of Agricultural Economics,vol.53. No. 1 pp. 53-62. https://doi.org/10.2307/3180297

Hazell PBR & Norton RD 1986, 'Mathematical Programming for Economic Analysis in Agriculture' Macmillan Publishing Company, New York p.400

Hardaker JB, Huirne RBM & Anderson JR, 1997, 'Coping with Risk in Agriculture' CAB International, Wallingford p. xi + 274.

Hardaker JB, Richardson JW, Lien G & Schumann KD 2004, 'Stochastic Efficiency Analysis with Risk Aversion Bounds: a Simplified Approach' Australian Journal of Agricultural Economics pp. 253-270.

Marler RT & Arora JS 2004, 'Survey of multi-objective optimization methods for engineering' Structural and Multidisciplinary Optimization, Volume 26, Issue 6, pp. 369–395. https://doi.org/10.1007/s00158-003-0368-6

Mendez MM & Sebastian RA & Iruretagoyena Osuna TM 1988, 'Cattle production systems by multiobjective programming for the tenth region of ChileAgricultural Systems' vol. 28, issue 2, pp. 141-157.

Ragsdale, CT 2007, 'Spreadsheet Modeling & Decision Analysis: A Prectical Introduction 1to Management Science' Fifth Edition, Thomson, p. 774.

Schneiderjans, MarcJ, Kwak N & Helmer MC 1982, 'An Application of Goal Programming to Resolve a Site Location Problem' Interfaces, vol.12. no.3 pp. 65-70.

Tóth J & Szenteleki K 1983, 'Mezőgazdasági vállalati célok elemzése kompromisszumos programozás segítségével' Sigma 3. Budapest. pp. 185-196.

Winston, WL 1997, 'Operations Research Applications and Algorithms' Wadswoth Publishing Company, p. 870

Zekri, S & Romero, C 1992, 'A methodology to assess the current situation in irrigated agriculture: An application to the village of Tauste (Spain)' Oxford Agrarian Studies, Volume 20, Issue 2. pp. 75-88. https://doi.org/10.1080/13600819208424050

Hungarian Association of Agricultural Informatics
European Federation for Information Technology in Agriculture, Food and the Environment

Journal of Agricultural Informatics. Vol. x, No. x
journal.magisz.org

Significance of food industry trademarks and product traceability from a customer perspective

István Füzesi[1], Péter Lengyel[2], János Felföldi[3]

I N F O
Received 2 Mar 2018
Accepted 7 May 2018
Available on-line 18 Jun 2018
Responsible Editor: M. Herdon

Keywords:
product trademark, food safety, traceability, food marketing.

A B S T R A C T

Food quality and safety related scandals have received considerable media coverage and have improved consumer awareness in Hungary and Europe. Consequently, the existence or lack of food safety may become a significant limiting factor which affects global trade. The number of consumers who require additional information on the origin of food products is becoming increasingly widespread. Our research seeks to provide an answer to the question whether confidence in the product manufacturer, some certificate (trademark) or detailed traceability information might well increase the acceptance and sales of a particular product. Moreover, the paper investigates what level of traceability information detail Hungarian consumers need, whether they intend to pay a higher price for traceability data beyond legislative requirements. Results show that consumers are mostly unaware of product traceability notions and they have limited insight into the product chain. As a result, they frequently identify product traceability with food safety; furthermore, the majority of consumers question the reliability of the information provided. The development, establishment and operation of a complex information system to provide adequate data for consumers require considerable investments and improvements, which are difficult - if not impossible - to realise from the approximately 10% additional charge considered as acceptable by consumers. In conclusion, such service can succeed if it has other, extra functionality as well.

1. Introduction

Food safety and the origin of food products raise an increasing number of various doubts among customers. Scandals related to food safety and the origin of food products sometimes overreacted by the media have shaken the confidence of customers in commercial food products in recent years (Lakner, Szabó & Hajdúné 2005). To some extent, as a consequence of this, new principles have been laid down in EU food regulations, where consumer health protection has become one of the essential principles. At present, the majority of food products are produced on a large scale, and due to globalisation, they reach a large number of consumers fast. Production, distribution and foodstuffs trade represent exceptionally complex business activities. Therefore the development of new procedures and methodologies are needed to safeguard food quality. However, as consumers have no overview of the circumstances of food production and applied technologies, confidence in manufacturers may provide indications in the selection of a specific product. Today, emphasis will be put on the issue of food safety in all sectors of the food industry (Ding, Xu, Li & Xie 2017). Over the past years, the average customer may have heard a lot about problems (many times in an extreme form) related to food products, different microbiological infections, physical and chemical contaminations. In an attempt to deal with emergency situations more efficiently and to promote consumer confidence in the credibility of products, these risk factors have resulted in the formulation

[1] István Füzesi
University of Debrecen
fuzesi.istvan@econ.unideb.hu
[2] Péter Lengyel
University of Debrecen
lengyel.peter@econ.unideb.hu
[3] János Felföldi
University of Debrecen
felfoldi.janos @econ.unideb.hu

of stricter regulations and measures (Cebeci, Guney & Alemdar 2008). Problems that can arise might be solved by compliance with the legislation (or by their enforcement) in this field, by the introduction of quality management systems, the traceability and identifiability of products beyond reasonable doubt (Loureiro & Umberger 2007). According to the definition of ISO 8402:1994: "traceability is the ability to trace the history, application or location of an entity using recorded identifications". Its primary objective is to distinguish and precisely identify the source of potential contamination, to efficiently recall the product and to remove them from the market. It is also essential that it can help avoid particular additives or ingredients of foodstuffs at the time of purchase, giving more excellent choices for customers among products. Therefore, movements among every unit are to be recorded for the participants of the food supply chain (Felföldi, Botos, Péntek, Szilágyi & Várallyai 2017), just like in the case of food production processes. The spread of problems due to the lack of food safety cannot be exclusively traced back to the increasing rate of world trade. Population density growth linked to urbanization, changes in lifestyle, modifications in consumption habits, increasing environmental pollution, the lengthening of the chain between food production and consumption, the fact that food production is becoming large-scale, the application of new, hazardous technologies, the increasing number of consumers in sensitive groups (e.g. elderly people) are all key factors (Farkas 2002; Bánáti 2004). The number of disturbing cases receiving considerable media coverage has improved consumer awareness. Therefore the existence or lack of food safety have become significant obstacles restricting global trade (Golan, Krissof & Kuchler 2004).

2. Literature review

The whole food chain should ensure compliance with the most restrictive quality standards and safety regulations. In all phases of the food chain from the purchase of raw materials through manufacturing, distribution and sales, quality requirements related to the given products, processes or methods of treatment should be met by retailers or supply units. Previously used, traditional methods are no longer satisfactory to solve these issues. Recently, as a result of various efforts (Szilágyi & Herdon 2013), the correlations of general quality issues aiming at enhancing the significance of food safety have become increasingly apparent than before. Consequently, progressing rules, regulations and production techniques are drawn up to produce safe food products (Schiefer 1997).

The development of an appropriate food traceability system implies multiple motivation factors: compliance with legislation (e.g. 178/2002 EU Decree), quality management issues, logistics optimisation, minimisation of risk (Stranieri, Cavaliere & Banterle 2017) and also marketing elements (Popp & Bánáti 2006). The present article, therefore, focuses exclusively on consumer attitude about buyers' requirement of access to traceability data and intention to pay the related costs. It can be increasingly observed that the system of consumer values is showing a positive change (Szakály, Pető, Popp & Jasák 2015), i.e. there is an increasing demand for health-conscious behaviour and quality food products (Szakály, Soós, Kovács & Polereczki 2016). A survey carried out in the United States of America revealed, 58% of respondents confirmed that they regarded meat products traceable to a certain processing plant, particularly to a given animal herd. It comes as no surprise that 74% identified traceability with quality, and asserted that traceable meat was equal with better meat.The same survey concluded that the overwhelming majority of shoppers (91%) included people who would spend more on traceable meat. Approximately 67% replied that they would purchase more if traceability were guaranteed (Cunningham 2008). As for another research on willingness to pay in the countries of the United States, Canada, Japan and the United Kingdom, traceability in itself is not motivating consumers to pay extra charges; however, further additional benefits may give enough motivation (Dickinson & Bailey 2005). The findings of a Chinese survey point out that despite consumers link traceability with the notion of food safety, for the most part, they are unwilling to pay more for it, or they would accept an extremely low (~10%) cost increase (Zhaoa, Qiaoa & Chena 2010). Industry experience suggests that this extra charge is insufficient for the operation of a system to ensure continuous and effective information that customers require. The area of Hungarian food industry has already seen similar surveys related to consumer perceptions of traceability. These indicate that Hungarian consumer confidence is higher in traceable food products; however, due to low information levels, the majority of consumers fail to recognise the main point of traceability. The

majority of consumers are unable to bring a concrete decision as to the guarantee or functional quality nature of traceability, and they connect it primarily with food safety. The identification of the so-called Traceable Resource Unit (Kim, Fox & Gruninger 1995), i.e. TRU is pivotal in the determination of the depth of traceability in the case of certain products on the producer side. TRU is an illustrative unit that enables segregated tracking in the supply chain. The determination of an optimal TRU size is a complex task to fulfil the criteria of keeping quantities of products at risk of potential removal, economic damage and image destruction to a minimum. Based on this approach, TRU should be as low as possible; however, the lower the core entity subject to tracking, the higher the number of units for segregated classification and data storage, making the solution even more costly. The identification of the level of tracking requires the examination of potential risk factors and their potential occurrence. It means the determination of an optimum point to meet market demands that causes the lowest possible damage if and when products are recalled, suits the capacities of the related business and the applied technology, feasible regarding organisation and technicality, and will prove to be the most economical solution. The importance of food safety is different for all food chain participants (Füzesi, Lengyel, Szilágyi & Ráthonyi 2016). As customers are the first to be interested in food safety, their expectation of traceability system is to ensure safety and trust (Lakner, Bánáti, Szabó & Kasza 2003). Traceability systems offer the following benefits for customers: protection of food safety by effective product recalls in emergency situations; the system provides options for allergy and food intolerance sufferers or those who struggle with their lifestyles' debilitating effects to avoid certain food ingredients by offering actual choices from food products derived from a variety of production methods (Herdon, Pető, Botos & Várallyai 2014). Full traceability of products can be achieved by using numerical and barcode systems as well as electronic and biological marking systems and a combination of appropriate levels. Identification requirements ensure continuity of tracking and reliability between independent partners (common language and information interoperability). In order to keep the tracking from the producer to the consumer, in each intermediate step, information about the product must be provided without modification, along with other related additional information. The most critical point of food traceability is the potential loss of comprehensive traceability. It may result from the loss or damage of identification labels, the interconnection of product cycles, the use of more than one identifier and the differences in the chain levels.

3. Methodology

To answer the questions, we used two sets of questionnaire data where the results were evaluated separately. However, they were used in one study and proved to be particularly useful for drawing conclusions relevant to our subject (Table 1) The first survey on the shopping habits of Hungarian consumers was carried out in 2016. This representative survey examined the information-gathering habits of Hungarian customers regarding food products (Csapóné & Péntek 2014). It was followed by another online questionnaire from 1 June to 15 July 2017, which investigated the relation between young (most likely under 40) Hungarian consumers representing massive customer demand (80% of respondents represent this age group) and product traceability information beyond legislative requirements.

One of the most common objectives of research is the segmentation of consumers or respondents, i.e. the formulation of groups that are relatively homogeneous, but at the same time well distinguishable from each other (heterogeneous). Group formation or clustering is an area of crucial significance in statistical methodology and a priority issue in practical applications as well (Simon 2006). Segmentation was carried out by cluster analysis, a multivariable statistical method often used in scientific research and practical applications, which summarises the methodology of clustering, i.e. group formation (Simon 2006). As cluster analysis is highly sensitive to outliers, as a first step, a simple chain method was applied to explore them and to eliminate them by the SPSS program. We evaluated the two variables involved in the investigation by the same scale, so standardisation was not required. As the number of groups to be formed was not predictable, the hierarchical architecture and in particular, the aggregate method (Ward's method, which aggregates those clusters, where the increase of the internal variance will be minimal) were chosen.

Table 1. Distribution of samples regarding critical background variables (N_I=505, N_{II}=159)

Name	Sample distribution	
	Survey I.	Survey II.
Gender-specific		
Male	46.9%	42.8%
Female	53.1%	57.2%
Qualifications		
Max. 8 years in primary school	9.1%	0.6%
Technical secondary, technical schools	36.0%	2.5%
School-leaving exam	40.8%	38.4%
Advanced level	14.1%	58.5%
By economic activity		
Active manual workers	39.4%	8.2%
Active non-manual workers	18.8%	45.3%
Young mothers on maternity or parental leave	2.4%	5.0%
Pensioners	30.7%	2.5%
Students	4.0%	36.5%
Unemployed people	3.0%	1.3%
Other inactive employees	2.2%	1.3%
On grounds of the buyer of foodstuff		
Primary buyer	66.3%	46.50%
Secondary buyer	32.9%	53.50%
Other	0.8%	
On a subjective basis about one's income		
Live(s) quite well, even save(s) money	3.6%	27%
Live well, (but can hardly save money)	32.5%	49.70%
Just enough to make ends meet, but saving is impossible	55.2%	18.90%
Sometimes not enough to live on	6.5%	3.80%
Faces problems of immediate survival	0.4%	0.60%

Source: Author's development, 2017

Following conversion, data were analyzed by SPSS software. Results obtained from crosstabs by SPSS show the number of real data from existing cases included in and omitted from the calculation. The evaluation of questionnaires was carried out by descriptive statistical methods (percentages, standard deviation, averages) and the frequency of the joint occurrence of the values of different variables was quantified in an actual contingency table. The $\chi 2$ statistical test was used to investigate the correlation between two variables. The null hypothesis of the test revealed independence, i.e. estimated and measured findings were equal. The significance level of Pearson's Chi-Square value table of results generated by SPSS demonstrates the independence of variables. Chi-square calculation at a significance level lower than 0.05 revealed that the variables were related. (Balogh, Bai, Popp, Huzsvai & Jobbágy 2015). The two ranks were compared by Spearman's correlation coefficients.

4. Results

In the early stages of our research, we investigated where customers could obtain their information on food products. It was required to enable the estimation of the efficiency of initiatives promoting the sharing of traceability data with customers. It is clear, however, that consumers tend to be curious about prices and ingredients. It is because the rate of "conscious" customers is steadily increasing today, similarly to the rate of food allergy and intolerance sufferers. The findings suggest that the number of information sources in addition to data specified by legislation is very limited. Consumers look mostly for data that they can collect directly, the conditions of production, the life-cycle of products will not be listed among factors to guide them in their choices. Customers could receive information by various identification numbers, barcodes or manufacturers' websites, but according to our survey findings, these are very rarely used (below 10%). It is based on some reasons. Our findings also indicate that relatively few products are labelled to provide additional information for customers. Therefore they can rarely collect information through these channels. Naturally, demand for extra information depends heavily on the type of food products, so it would be wrong to deal with all product cycles and their complexity on the same footing.

Our survey addressed the significance of processing conditions for the production of food products purchased by customers and also end-to-end traceability during their sales process. The answers were measured on a scale of 1-5, where 1 represented the lowest approval rating and 5 the highest one. Our research investigated the milk, meat, vegetable and fruit product cycles separately, but our findings showed no significant differences among the product groups. The proportion of consumers who strongly agreed was determined by adding 4 and 5. The proportion of these respondents was above 60% in the case of all the three product groups. Personal interviews, however, highlighted the lack of awareness of customers about the meaning of "from fork to table" related to food traceability. First, their awareness should be improved about the mandatory nature of the registration of traceability data, and feasibility of effective product recalls in the framework of European Union and Hungarian legislation. Among others, the General Food Law Regulation 178/2002/EC of the European Parliament and Council make provision for this issue; however, it does not exclusively cover the obligation to make these data available for customers. Second, TRUs of different volumes are used for traceability data collection, which means that only information related to the potential sources of food products can be registered instead of the exact list of sources they include. Spearman's rank correlation calculation revealed a significant correlation ($p<0.05$) between consumers' demand for traceability data, their level of health-consciousness and monthly income. Notably, a similarly strong correlation exists between the two explanatory variables, i.e. those with high incomes are willing to spend more money on healthier food products.

Demand for food trademarks is even higher. 87% of consumers replied that trademarks certifying the origin of liquid milk (fresh, not long-life), fresh poultry meat, fresh pork, fruits and vegetables would be necessary. Trademarks identify the product origin thus it gains a certain degree of specificity, and raises the opportunity for customers to differentiate between specific products; hence trademarks play a key role in food marketing. Consequently, the primary considerations in product selection will be previous shopping experience, confidence in manufacturers and certified trademarks. Interestingly, customers who think they are health-conscious, do not consider the use of trademarks certifying the origin of food products necessary; the obtained data failed to reveal any correlations. On the contrary, a low, but detectable correlation emerges between household incomes and demand for trademarks: the higher the standard of living, the increased the interest in them.

Next, consumers were broken down into homogeneous groups by cluster analysis by two relevant variables (Table 2). The two variables included demand for traceability data about health consciousness and meat products (as we have seen, there was no deviation among product groups from this point of view). Our investigations resulted in the formation of three clusters. The first group included "neutral" consumers, the second those who find traceability data essential but do not regard themselves to be health-conscious, and the third those who considered both factors highly significant. On the basis of the findings, the majority of consumers could be divided into the first cluster.

Table 2. Distribution and preferences of consumer groups

Ward Method		How important is end-to-end traceability related to the conditions of processing and sales in meat production for you?	To what extent are you health-conscious?
1	Mean	2.94	3.35
	N	242	242
	Std. Deviation	1.024	,650
2	Mean	4.74	2.52
	N	107	107
	Std. Deviation	,574	,711
3	Mean	4.82	4.29
	N	149	149
	Std. Deviation	,518	,454
Total	Mean	3.89	3.46
	N	497	497
	Std. Deviation	1.230	,880

Source: Author's development, 2017

Of the investigations on the correlations of information capture regarding the clusters and food products, correlations of trademarks and markings (Table 3) were the most relevant in our research topic. Analysis on the relationship of clusters and trademarks suggests that consumer groups correlate demonstrably with demand for trademarks. The first group consisting of the majority of customers showed little interest in trademarks on food products.

Table 3. Relationship between consumer groups and trademarks

If you want to get information on food products, what sources do you use? Trademark, marking on the label * Ward Method Crosstabulation

			Ward Method			Total
			1	2	3	
If you want to get information on food products, what sources do you use? Trademark, marking on the label	**Yes**	Count	74	50	81	205
		% within Ward Method	30.6%	47.2%	54.4%	41.2%
		Adjusted Residual	-4.7	1.4	3.9	
	No	Count	168	56	68	292
		% within Ward Method	69.4%	52.8%	45.6%	58.8%
		Adjusted Residual	4.7	-1.4	-3.9	
Total		Count	242	106	149	497
		% within Ward Method	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	23.476[a]	2	,000
Likelihood Ratio	23.658	2	,000
Linear-by-Linear Association	22.680	1	,000
N of Valid Cases	497		

a. 0 cells (,0%) have expected count is less than 5. The minimum expected count is 43.72.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,217	,000
	Cramer's V	,217	,000
	Contingency Coefficient	,212	,000
N of Valid Cases		497	

Source: Author's development, 2017

We examined the correlations of consumers classified in specific clusters and demographic properties (age group, gender, qualifications, marital status, number of dependents, residence, region, and income, the primary buyer of food products, environmental awareness, and preference of Hungarian products). We drew the conclusion that inclusion in certain clusters was affected by the gender of customers, their health-consciousness, commitment to Hungarian products and the person who was the primary buyer of food products in households. These findings enabled the characterisation of particular groups (Table 4). The analysis did not cover the criteria of environmental awareness, as it was closely linked to the quality criteria of health consciousness.

Table 4. Characterization of consumers divided into clusters

		Cluster 1.	**Cluster 2.**	**Cluster 3.**
Variables in cluster processes	Demand for traceability data	Neutral	Yes	Yes
	Health consciousness	Neutral	No	Yes
Variables included exclusively in characterisation	No	In equal number	In equal number	Women
	Primary buyers of food products	Mostly yes	Mostly yes	Yes
	Commitment to Hungarian products	Mostly yes	No	Yes
Name		*Neutral*	*Interested*	*Careful housewives*

Source: Author's development, 2017

Our findings of the characteristics indicate that the first group consists of customers who showed little interest in the origin of food products and health-conscious lifestyle. Customers involved in the second cluster attach importance to food safety, although they would not refer to themselves as

conscious consumers. The third cluster indicates those female buyers who care about their families and place a high emphasis on safety and health (>81%). Responses to our questionnaire questions about food safety reveal that Hungarian consumers regard the level of food safety in Hungary appropriate (82.4% of them gave medium and good ratings). We also investigated whether authority control or product traceability guaranteed better food safety to customers. On the basis of the findings, it can be concluded that both factors influence the feeling of security significantly ($p<0.05$), and Hungarian consumers have confidence in the work of Hungarian authorities and food traceability.

The next part of our research attempted to answer questions about customers' considerations regarding those products with traceability data providing information in addition to legislative requirements; furthermore, whether extra traceability exerts an influence on product price and quality. Our results demonstrate that in the opinion of consumers, food traceability affects quality and food safety the same way. Therefore, they claim that products that contain these data are sold at higher prices than "traditional products". The proportion of the potential customers of products that contain traceability data beyond legislative requirements was compared to data from the Chinese study mentioned in the literature survey (Table 5). Not surprisingly, it can be seen that the majority of consumers would purchase products that contain traceability data beyond legislative requirements at a sale price equal to traditional products. It is especially interesting why the remaining percentage of customers would bring a different decision as no harm came to them if they received additional information.

Table 5. The proportion of customers purchasing products containing data beyond legislative requirements.

	Peking	*Xianyang*	*Hungary*
At the same price as traditional products	93%	97%	89%
At a price higher than that of traditional products	55%	50%	64%
Rate of acceptable extra charge	11%	11%	1-10%

Source: Zhao et al. 2010 and authors' survey, 2017

However, just over half of customers would choose traceable products at prices higher than traditional products; according to the survey, this proportion is higher in Hungary (64%). On the contrary, if the rate of the acceptable extra charge is lower, customer demand will only accept 1% - 10%. Our investigations explored a precise correspondence between the demand for extra traceability data and the level of household income and health-conscious behaviour ($p<0.05$). Our study has found that the implementation of the development, improvement and operation of a traceability system mentioned in the introduction, at such an extra charge seems to be very doubtful. An additional criterion provides that customers should demand these data in the long-term regularly and should pay for it; however, our survey suggests that merely a low percentage of customers would be willing to act accordingly.

4. Conclusion

Food product traceability has three motivating factors. The first is the consideration of the government, i.e. effective traceability in public health and for problematic products. All products shall meet these requirements. Food producers are motivated to implement agile traceability by the compliance of domestic and EU legal requirements, standards for suppliers, commercial quality management systems (e.g. IFS, BRC) and by effective product recalls (with fewer costs). Our research asked whether it might be worthwhile operating a system that provides additional information beyond these factors. The enterprises in our case studies have claimed that due to the low number of product

recalls, investment in such projects is worthwhile if extra costs are financed by customers. Then our research attempted to find out in what conditions customers would be willing to pay these costs. However, the implementation of these information services is doubtful in various aspects. First, several investigations have proved that most customers are not aware of the essential nature of traceability in all cases, and their overview of the product chain is limited. As a result, they frequently identify product traceability with food safety. Furthermore, the majority of consumers doubt the reliability of the information provided; their demand for these data should be continuous and regular to meet an additional critical criterion. However, there is little chance for average consumers to query about data of the same product type permanently.The rate of an acceptable extra charge is also questionable, if interpretable, as more than half of our respondents would never find the consequent price rise acceptable. The development, establishment and operation of a complex information system to provide consumers with adequate data would require massive investments, which are very difficult or impossible to implement from the 10% extra charge acceptable for customers. In conclusion, such service can succeed if they provide additional functionality as well (organic, traditionally produced food product, etc.) Our research shows that trademarks and confidence in producers are of greater importance than the accurate and continuous availability of traceability data.

References

Bánáti, D 2004, 'Élelmiszer-biztonság az élelmiszeriparban.' (Food safety in the food industry.) Kopint-Datorg, Budapest, pp. 1-46.

Balogh, P, Bai, A, Popp, J, Huzsvai, L, & Jobbágy, P 2015, 'Internet-orientated Hungarian car drivers' knowledge and attitudes towards biofuels' *Renewable & Sustainable Energy Reviews* 48: pp. 17-26. https://doi.org/10.1016/j.rser.2015.03.045

Cebeci, Z, Guney, I, & Alemdar, T 2008, 'Designing a Conceptual Production Focused and Learning Oriented Food Traceability System.' In Proceedings of the 4th International Conference on Information and Communication Technologies in Bio and Earth Sciences, (Ed. T. Tsiligiridis), 18-20 Sep 2008,Agric. Univ. Of Athens, Greece. pp. 206-213.

Csapóné Riskó, T, & Péntek, Á 2014, 'A tudatos fogyasztás megjelenési formái.' (Forms of Conscious Consumption.) *Táplálkozásmarketing* 1:(1-2) X. pp. 73-80.

Cunningham, P 2008, 'Using DNA Traceability to Track Meat and Ensure Safety.' *Genetic Engineering & Biotechnology News*, Vol. 28, No. 8,

Dickinson, D, L, & Bailey, D 2005, 'Experimental Evidence on Willingness to Pay for Red Meat Traceability in the United States, Canada, the United Kingdom, and Japan.' *Journal Of Agricultural & Applied Economics* 37/3 pp. 537-548.

Ding, J, Xu, H, Li, P, & Xie, R 2017, 'Design and implementation of food safety traceability system based on RFID technology.' *Advances in Intelligent Systems and Computing*, Volume 611, pp. 657-666. https://doi.org/10.1007/978-3-319-61566-0_61

Farkas, J 2002, 'Élelmiszer-biztonság: globális gondok – javítási törekvések.' (Food security: global concerns - improvement aspirations.) Magyar Tudomány 2002/12. Budapest, 1608. p.

Felföldi, J, Botos, Sz, Péntek, Á, Szilágyi, R, & Várallyai, L 2017, 'Studying the ict management of agrifood sector on supply chain level – the first stage: analysis of agricultural ict usage' In: Anastasios Karasavvoglou, Persefoni Polychronidou (ed.) Agricultural Sector Issues in the European Periphery: Productivity, Export and Development Challenges. Wilmington (DE): Vernon Press, 2017. pp. 35-50.

Füzesi, I, Lengyel, P, Szilágyi, R, & Ráthonyi, G 2016, 'Application of EDI Technologies in the Food Supply Chains', *Journal of Ecoagrotourism* 12:(1 (32)) pp. 69-77.

Golan, E, Krissof, B, & Kuchler, F 2004, 'Food Traceability: One Ingredient in a Safe and Efficient Food Supply'. Amber Waves April, 2004. Economic Research Service/USDA

Herdon, M, Pető, K, Botos, Sz, & Várallyai, L 2014, 'The Rules of ICT in Regional Development Related to the Agri-Food and Tourism Sectors.' *Journal of Ecoagritourism* 10(1) pp. 174-181.

Kim, H.M, Fox, M,S & Gruninger M 1995, 'An Ontology of Quality for Enterpise Modeling, Enabling Technologies' *Infrastructure for Collaborative Enterprises, 1995., Proceedings of the Fourth Workshop Berkeley Springs*, WV, pp. 105-116.

Lakner, Z, Bánáti, D, Szabó, E, & Kasza, Gy 2003, 'A magyar fogyasztó és az élelmiszerek biztonsága.' (The Hungarian consumer and the safety of food.) *A Hús*, 2003/1. 49-55. pp.

Lakner, Z, Szabó, E, & Hajdú, I.-né 2005 'The 2004 paprika scandal: anatomy of a food safety problem.' *Studies in Agricultural Economies*. Budapest, Agricultural Economics Research Institute, pp. 67-82.

Loureiro, M, L, & Umberger, W, J 2007, 'A choice experiment model for beef: What US consumer responses tell us about relative preferences for food safety, country-of-origin labeling and traceability.' *Food Policy* 32 (2007) pp. 496–514. https://doi.org/10.1016/j.foodpol.2006.11.006

Popp, J, & Bánáti, D 2006, 'Élelmiszer-biztonság a nemzetközi kereskedelem tükrében.' (Food safety in the light of international trade.) Agrárgazdasági Kutató Intézet, Budapest, 141. p.

Schiefer, G 1997, 'Quality Management and Process Improvement – The Challenge', In: Schiefer G., Helbing R. (eds.), Quality Management and Process Improvement for Competitive Advantage in Agriculture and Food (Volume I.), Dept. of Agricultural Economics, University of Bonn, 509. p.

Simon, J 2006, 'A klaszterelemzés alkalmazási lehetőségei a marketingkutatásban.' (Application possibilities for cluster analysis in marketing research.) *Statisztikai Szemle* 84. évfolyam 7. szám pp. 627-651.

Stranieri, S, Cavaliere, A, & Banterle, A 2017, 'Do motivations affect different voluntary traceability schemes? An empirical analysis among food manufacturers.' *Food Control* Volume 80, 1 October 2017, pp. 187-196. https://doi.org/10.1016/j.foodcont.2017.04.047

Szakály, Z, Pető, K, Popp, J, & Jasák, H. 2015, 'A fenntartható fogyasztás iránt elkötelezett fogyasztói csoport, a LOHAS szegmens jellemzői.' (A consumer group dedicated to sustainable consumption, characteristics of the LOHAS segment.) *Táplálkozásmarketing* 2:(1) pp. 3-9.

Szakály, Z, Soós, M, Kovács, S, & Polereczki, Zs 2016, 'A felárfizetési hajlandóságot befolyásoló tényezők elemzése funkcionális élelmiszereknél.' (Analysis of factors influencing the willingness to pay on functional food.) *Gazdálkodás* 60:(4) pp. 305-321.

Szilágyi, R, & Herdon, M 2013, 'Augmented Reality (AR) Applications in Agriculture.', In: Zacharoula Andreopoulou, Vagis Samathrakis, Soulla Louca, Maro Vlachopoulou (ed.) E-Innovation for Sustainable Development of Rural Resources During Global Economic Crisis. Hershey (PA): IGI Global, Information Science Reference, 2013. pp. 65-79.

Zhaoa, R, Qiaoa, J, & Chena, Y 2010, 'Influencing factors of consumer willingness-to-buy traceable foods: An analysis of survey data from two Chinese cities.' *Agriculture and Agricultural Science Procedia* 1 pp. 334–343. https://doi.org/10.1016/j.aaspro.2010.09.042