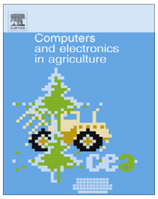
Computers and Electronics in Agriculture 120 (2016) 72–78

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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



Original papers

Analyzing the behavior dynamics of grain price indexes using Tucker tensor decomposition and spatio-temporal trajectories

F.E. Correa [a,*], M.D.B. Oliveira [b], J. Gama [b], P.L.P. Corrêa [a], J. Rady [a]

[a] *Department of Computer Engineering – Polytechnic School – University of São Paulo, PO Box 61548, São Paulo, SP 05424-970, Brazil*
[b] *Laboratory of Artificial Intelligence and Decision Support – INESC TEC, University of Porto, R. Doutor Roberto Frias 378, 4200-465 Porto, Portugal*

ARTICLE INFO

Article history:
Received 13 October 2014
Received in revised form 16 November 2015
Accepted 22 November 2015
Available online 7 December 2015

Keywords:
Data mining
Tucker decomposition
Spatio-temporal trajectories
Grain market

ABSTRACT

Agribusiness is an activity that generates huge amounts of temporal data. There are research centers that collect, store and create indexes of agricultural activities, providing multidimensional time series composed by years of data. In this paper, we are interested in studying the behavior of these time series, especially in what regards the evolution of agricultural price indexes over the years. We explore data mining techniques tailored to analyze temporal data, aiming to generate spatio-temporal trajectories of grains price indexes for six years of data. We propose the use of Tucker decomposition to both analyze the temporal patterns of these price indexes and map trajectories that represent their behavior over time in a concise and representative low-dimensional subspace. The case study presents an application of this methodology to real databases of price indexes of corn and soybeans in Brazil and the United States.

1. Introduction

The agricultural commodities are very important to economies of several countries, especially Brazil, where these assets account for 7.3% of the Gross National Product – GNP. Moreover, agricultural activities are the backbone of most economic systems, in the sense they represent an important source of raw materials to other industries (e.g., cotton, sugar) and provide many employment opportunities for the labor force (IBGE).

Two of the most important agricultural products in the Brazilian economy are corn and soybean. These products belong to the family of grains. For example, Brazil exportation of soybeans grains was 32 M. tons, representing an estimated figure of 17.5 billion dollars in 2012 (ALICEWEB).

Despite the great amount of money involved, we do not have, in agribusiness activities, accurate information for the whole process. Therefore, research centers in Brazil, such as the Center for Advanced Studies on Applied Economics – CEPEA, collect and provide price indexes of these commodities (Correa, 2009).

Studies to understand the temporal trajectory of a variable, such as prices for products like soybean and corn, provide the market players with strategic information regarding the international market transaction behavior over the last years. Considering that the models were applied on real data, it is possible to update these models with new collected data and use them to infer or predict if some events will continue to happen. Moreover, we can mention some other benefits that arise from the international market analysis. For example, it is possible to observe that the trajectory of the Chicago stock market prices is the base for price indexes over the Brazil internal prices. As a result, further research could try to measure what is the impact of some public policies, e.g. American incentives for corn producers, by adding such information on new models and simulations (Aruga, 2014; Rosa et al., 2014).

In order to analyze agro economic data, it is necessary to join several databases of distinct types and subjects (Plant, 2012). Databases with this kind of information are usually multidimensional, i.e., they have more than two dimensions (variables that affect their behavior).

Examples of multidimensional data are common in agriculture. Usually, the products are negotiated in different types of markets, e.g., domestic market and stock market. Further, there are a variety of products, like corn, soybeans grain and meal. Moreover, these multidimensional data are temporally ordered, i.e., they are time series collected and stored over several years (King, 2010).

There are data mining techniques able to deal with multidimensional and temporal data. In this research, our aim is to explore two of those techniques in order to provide a methodology for the analysis of agro economic data. The main techniques and framework used were Tucker decomposition (Tucker, 1966; Oliveira and Gama, 2012) and spatio-temporal trajectories (Oliveira and

[*] Corresponding author.
E-mail addresses: fecorrea@usp.br (F.E. Correa), mdbo@inescporto.pt (M.D.B. Oliveira), jgama@fep.up.pt (J. Gama), pedro.correa@poli.usp.br (P.L.P. Corrêa), jorge.rady@usp.br (J. Rady).

http://dx.doi.org/10.1016/j.compag.2015.11.011
0168-1699/© 2015 Elsevier B.V. All rights reserved.

Gama, 2013). Some complementary statistical methods were used, namely, the correlation matrix (Kazmier, 2004), ANOVA and the sliding window model (Datar et al., 2002). In addition, clusters analysis for pattern mining applied on spatio-temporal data were reviewed to consider by future implementation (Patel, 2005; Xiao, 2014).

Using a methodology based on the aforementioned data mining techniques, the idea is to understand the evolution of the time series of Grains price indexes over a time span of six years. In the case study presented it was used a real-world time series. Our main contribution is to propose a process methodology to identify, summarize and highlight past events and provide analysis methods to deal with multidimensional datasets.

This paper is organized as follows: In Section 2, we introduce the main concepts about the Tucker decomposition and data mining techniques for analyzing temporal data. After providing the background, we detail the proposed methodology in Section 3. The next section presents the application of the proposed methodology to multidimensional time series of grains price indexes. This paper ends with the related work, conclusions and suggestions for further research.

2. Tucker decomposition

Tucker decomposition is an unsupervised multiway data analysis method that is quite useful for data cleaning, data compression and visualization of the main structures of data in low-dimensional spaces. Tucker (1966) devised this method in order to extend the well-known PCA (Principal Component Analysis) to higher-order data representations, such as tensors. We can straightforward define a tensor as an extension of a matrix to three or more dimensions, or as an N-way data array, where N is the *order* of the tensor. The Data Mart analyzed in this paper can be arranged into a three-order tensor, by incorporating the temporal dimension. We resort to three-order tensors, instead of matrices, in order to explicitly account for the time dimension and, thus, avoid loss of information in the modeling process. The *order*, *ways* or *modes* of a tensor are synonyms and refer to the number of dimensions (in our case, we have three dimensions: products, market and time). For this specific type of tensors or, in other words, N-way data arrays for $N = 3$, the most appropriate Tucker decomposition model is the so-called *Tucker3 tensor decomposition* (Kolda and Bader, 2009), which performs the reduction of data in all three modes of the tensor.

The basic idea of the Tucker3 decomposition is to find a set of matrices (known as the *component matrices*) and a small tensor (known as the *core tensor*) that, in general, have less dimensionality than the original tensor, but are able to reconstruct the most important information contained in data.

The Tucker3 model can be formulated as the factorization of the original three-order tensor χ, such that

$$\chi_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} g_{pqr} a_{ip} b_{jq} c_{kr}$$

for $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $k = 1, \ldots, K$. Here, the coefficients a_{ip}, b_{jq} and c_{kr} represent the entries of the *component matrices* $\boldsymbol{A} \in \mathbb{R}^{I\times P}$, $\boldsymbol{B} \in \mathbb{R}^{J\times Q}$ and $\boldsymbol{C} \in \mathbb{R}^{K\times R}$. In turn, the coefficient g_{pqr} represents the entry of the so-called *core tensor* $\mathcal{G} \in \mathbb{R}^{P\times Q\times R}$. The number of entities in each mode are represented by letters I, J and K. The number of components (i.e., the number of columns of the matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$) in the first, second and third mode of the tensor are represented by letters P, Q and R, respectively. This decomposition is illustrated in Fig. 1.

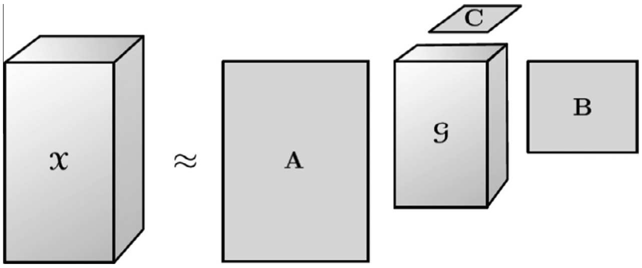


Fig. 1. The basic Tucker3 tensor decomposition (Kolda and Bader, 2009).

Tucker suggested that the *core tensor* $\mathcal{G}$ can be interpreted as describing the latent structure in the data and the *component matrices* ($\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$) as mapping this structure to give the observed data (Tucker, 1966). The *core tensor* can also be interpreted as a generalization of the eigenvalues of the SVD (Singular Value Decomposition) (Skillicorn, 2007). Detailed information about the Tucker3 technique can be found in Tucker (1966), Kolda and Bader (2009) and Kiers and Mechelen (2001).

This technique can be used to detect abnormal events and important milestones in the agribusiness data, by means of the projection of spatio-temporal trajectories in Tucker3 bi-dimensional subspaces. Spatio-temporal trajectories visually represent the movement of a given object in a plane. They can be formally defined as a function from the temporal dimension $I \subseteq \mathbb{R}$ to the geographical space $\mathbb{R}^2$ (*i.e.*, the 2D, or bi-dimensional, space) (Kiers and Mechelen, 2001). At each time point, the object occupies a given position in the 2D space. Each position is recorded in terms of (x, y) coordinates, which represent latent concepts, and associated with the corresponding time stamp. The temporally ordered sequence of an object's positions defines the trajectory of this object, which are often represented as (x, y, t) triples:

$$T = \{(x_1, y_1, t_1), (x_2, y_2, t_2), \ldots, (x_k, y_k, t_k)\}$$
$$\text{where } x_i, y_i, t_i \in \mathbb{R} \ (i = 1, \ldots, k) \text{ and } t_1 < t_2 < \ldots < t_k$$

These trajectories are graphically represented by a line that connects the coordinates of each position to the object's movement. The goal toward the use of spatio-temporal trajectories is the representation of time series in a way that is efficient to analyze. The analysis of trajectories allows us not only to understand the dynamics of an object's behavior (*e.g.*, the evolution of corn indexes with respect to a set of agro economic indicators) but also to understand large quantities of information in a concise way.

3. Methodology

3.1. Grain dataset specification

In order to make possible the analysis of the multidimensional databases, we had to do some procedures on the data, in order to retrieve time-ordered data in a format that can be used to generate the spatio-temporal trajectories. The multidimensional dataset was split into 3 dimensions, or *modes.* One of these dimensions is time. The time dimension can have several granularities. In our case, the time unit selected is the month. To obtain monthly data from years 2007 to 2012, six datasets were created, one for each year. Considering this division it was possible to label the trajectories' variables per year in the plot.

Each dataset that results from the application of the aforementioned technique will be called a *data cube* (i.e., a three-order tensor). The data cube used in this paper has 3 dimensions or modes, namely: products, market and time. The entities that belong to the products dimension are the collected price indexes for the following grains: corn, soybeans grain, soybeans meal and soybeans oil. These price indexes were obtained for several types of markets, namely, the Brazilian domestic market, the US Chicago Stock market and the exportation figures of Brazil and the United States. These entities are associated to the second dimension of the data cube (i.e., the market dimension). The last dimension is time and the corresponding entities are the months associated with each year.

Using this approach, six data cubes were created. These data cubes refer to each one of the analyzed years (2007–2012) and were arranged into three-order tensors. The top panel of Fig. 2 represents the three-order tensor, comprising three dimensions and the corresponding entities, considered in this study. The bottom panel of Fig. 2 depicts the tensors generated for each year. Each tensor has the following modes, or dimensions: time, measured in months (row-entities); type of market (column-entities), and products (fiber-entities).

3.2. Application and analysis of the Tucker decomposition

After modeling the multidimensional agribusiness data using three-order tensors, patterns were extracted from each one of the six tensors by applying the Tucker3 model, using procedure develop in R software, adapted from studies from Oliveira and Gama (2012) and Pebesma (2012). This model summarizes the information contained in the tensor in a set of components. These components represent, in a concise way, the main variability of the original datasets, thus being able to uncover the most important patterns in the data.

To apply the Tucker3 model, it was necessary to define what combination of components adjusts better for each tensor. This combination of components is basically the number of patterns to extract from each dimension of the tensor. These components summarize the entities in each dimension. One way to select the appropriate number of components is to use a procedure known as the Scree Plot.

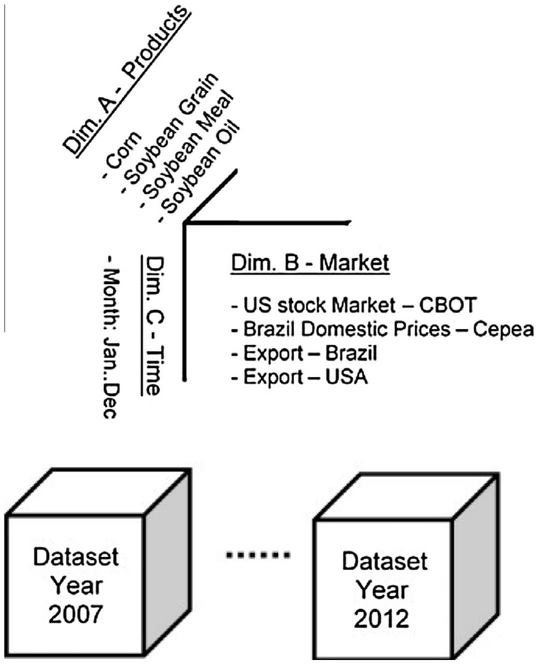


Fig. 2. Example of a three-order tensor for each year of the grain dataset. Each tensor has the following modes, or dimensions: time, measured in months (row-entities); type of market (column-entities), and products (fiber-entities).

Scree Plot simulates all possible models (P, Q, R) that can be used by Tucker3, i.e., all combination of components for each dimension. The selection of the appropriate model is performed by balancing parsimony, or model dimensionality (i.e., the fewer components in each dimension), and fit (i.e., the amount of the total variance captured by the model). For guiding the choice of the number of components retained in each dimension, the threshold for the model fit used in this paper was 85%. It is important to emphasize that this threshold was chosen based on the knowledge of the domain. Therefore, the appropriate choice of this threshold depends on the application domain and its complexity. This is important because, the higher the number of components, the higher the complexity of the results and the higher the processing time. Based on this threshold, we consider only the combination of components explaining at least 85% of the total variance, and we disregard the models with lower fit. After applying this threshold, we select the most parsimonious model from the available options, i.e., the model with lower dimensionality. Since we aim at obtaining a concise representation of the original three-order tensors, the number of components extracted for each dimension should be lower than the original number of entities (i.e., $P < I$, $Q < J$ and $R < K$).

In Fig. 3 it is presented the Scree Plot of year 2007. This means that we only consider those models that are able to explain at least 85% of the variability contained on the original three-order tensor. From Fig. 3, it can be ascertained that the most appropriate combination of components (P, Q, R) is (2, 3, 3), which explains 90% of the total variance for dimensions A, B and C, respectively.

The analysis of the scree plot is repeated for the data cubes corresponding to years 2008, 2009, 2010, 2011 and 2012. In Fig. 4, we depict the Scree Plot of year 2008. Using the same procedure, the selected model is (3, 3, 3). In this plot it is possible to see that models comprising a higher number of components, such as (4, 2, 3) explain less data variability than the model we chose, which is more parsimonious. This means that an increase in the model complexity does not always translate into a better model fit (i.e., a higher explained variance). Since the results of the intersections variability evolve in all the dimensions together, this can decrease the percentage of explained variation, or the complexity of the calculus.

It is possible to observe in both Figs. 3 and 4 that the upper bound of possible models is given by the combination (4, 4, 4). In addition, almost 40% of component C corresponds to 100% of variability of the dataset. This combination (4, 4, 4) is not chosen due to the two criteria explained before: first, the model complexity (too many components, or patterns, to analyze) and, second, the required processing time.

After analyzing the Scree Plot for all cubes/tensors, we selected the models (P, Q, R) presented in Table 1 for application of the Tucker3 decomposition.

4. Results – grain time series – a case study using trajectories analyses

The final step of the methodology is the generation of the spatio-temporal trajectories based on the output of the selected Tucker3 models. The idea was to use the Tucker3 score as a weight of the entities in each dimension of the three-order tensor.

To better understand the procedure to obtain the trajectories, we will take dimension A (products) as an example. Dimension A has 4 entities: corn, soybean, meal and oil. The application of the Tucker3 model (2, 3, 3) for year 2007, produced a set of scores that summarizes the four entities into three factors, or components. These scores are reported in Table 2. The next step is the generation of spatio-temporal trajectories. These trajectories are

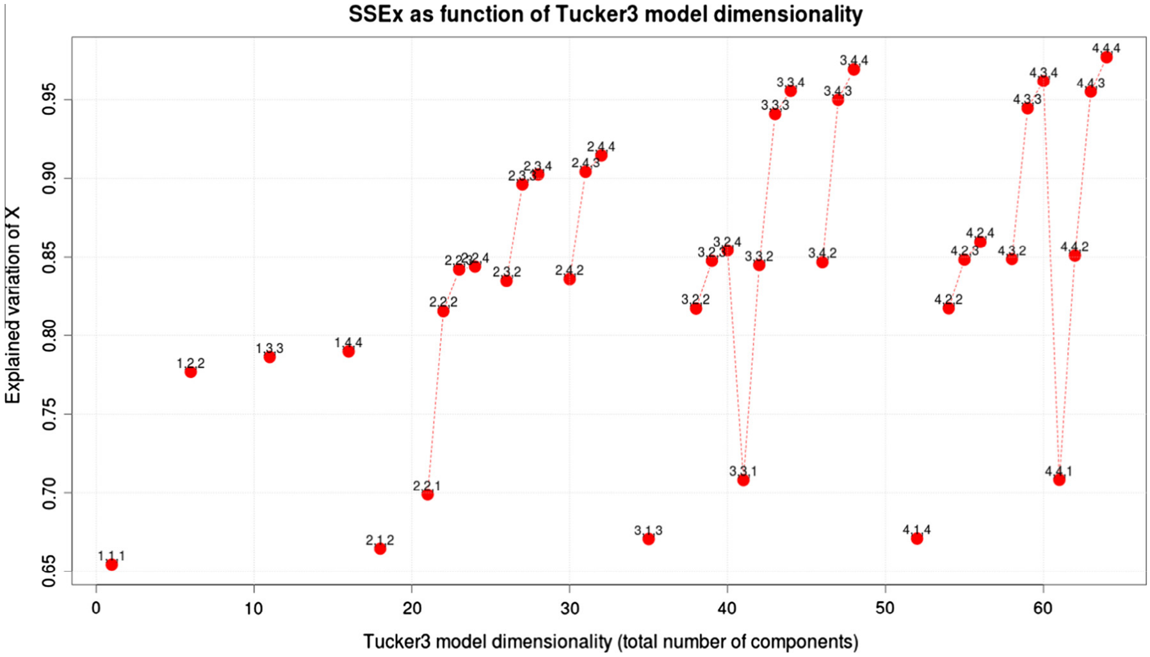


Fig. 3. Scree Plot of year 2007.

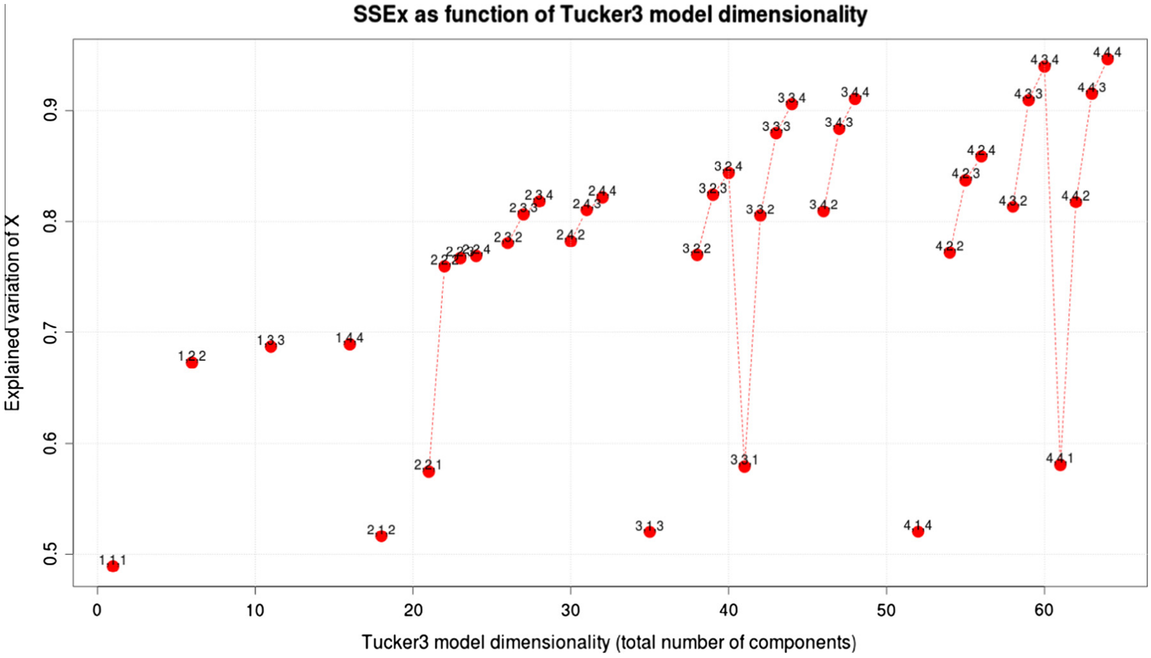


Fig. 4. Scree Plot of year 2008.

Table 1
Models selected for each cube/tensor.

Year	Model (*P*, *Q*, *R*)	Proportion of explained variability (%)
2007	2, 3, 3	90
2008	3, 3, 3	88
2009	3, 3, 3	85
2010	3, 3, 3	88
2011	3, 3, 3	86
2012	2, 3, 3	89

Table 2
Tucker3 model scores for dimension *A* and year 2007.

Scores dimension *A* – year 2007			
Variable	1° Factor	2° Factor	3° Factor
Corn	−0.41	0.60	−0.66
Soybean grain	−0.42	−0.61	−0.14
Soybean meal	−0.37	−0.45	−0.37
Soybean oil	−0.72	0.23	0.64

projected in a bi-dimensional Tucker subspace, i.e., a 2D plane where the *x* axis corresponds to the first component of dimension *A* and the *y* axis corresponds to the second component of the same dimension. The first and second factors of each dimension are selected because we intend to generate a representative bi-dimensional space that explains the greatest amount of variability.

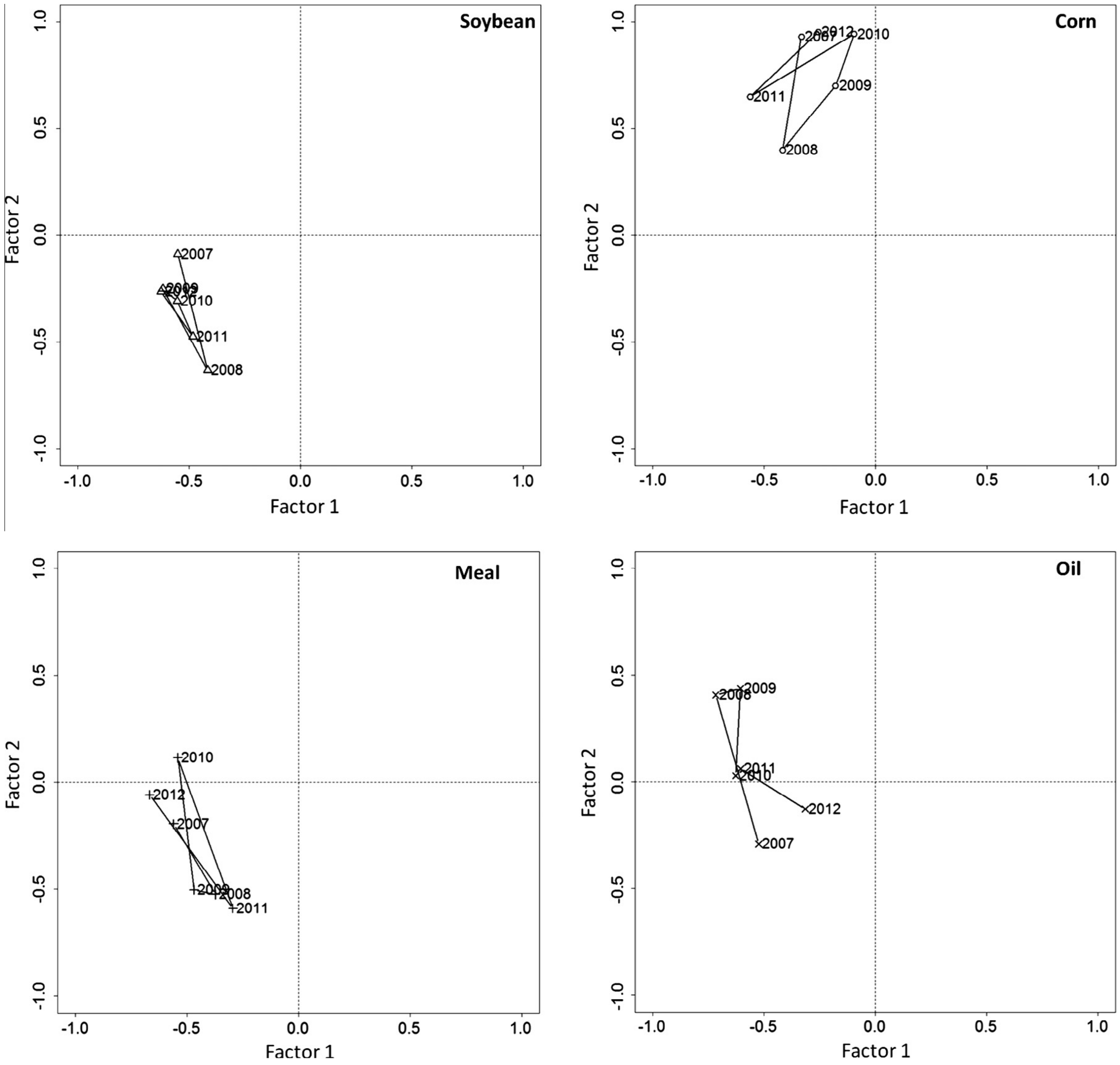


Fig. 5. Four trajectories of dimension *A* – products.

Since the two first components explain a larger proportion of the variability, we exclude the third factor from the analysis.

To illustrate the process of generating a spatio-temporal trajectory, consider the first entity of dimension *A*: the product corn. For each year, we plot the corresponding position $p_i = (x_i, y_i)$. The first point, corresponding to year 2007, is $p_1 = (-0.41, 0.60)$. These coordinates are the scores obtained by the entity "corn" in the first and the second components yielded by the application of the Tucker3 model to the three-order tensor of year 2007 (see Table 2). The positions of the entity corn in the following years are obtained using the same procedure.

After obtaining all the positions associated with the movement of corn, the trajectories are drawn by connecting these points in the 2D space. The resulting spatio-temporal trajectories for corn and the remaining entities of dimension *A* are depicted in Fig. 5. Each point is labeled by year. The analysis of these trajectories allow us to understand the behavior of the products in the past.

The first thing we can highlight in Fig. 5 is that the corn trajectory is located in the upper quadrant of the 2D space, whereas the soybean grain and the corresponding derivatives are located roughly on the same quadrant. This result shows that, since corn and soybean crops compete for planting area, their prices are exclusively distant from each other.

For the dimension *B* – market, the results presented in Fig. 6 were divided by showing each trajectory in an exclusive plot. It is possible to see that the trajectories of CBOT and CEPEA are closer. As explained earlier, the CBOT prices are usually mandatory in grains commercialization, and the CEPEA trajectory reflects this. Another interesting aspect is that the exportation in both countries has an opposite trajectory design.

Considering dimension *C* – time (measured in months), the variation of all months does not allow to make any conclusion, once there are 12 variables and the points are spread for all quarters of the 2D plane.

Although we do not explore the depth of dimension *C*, further analysis will be conducted in future work aiming to provide integrated scenarios to analyze the impact that a trajectory of one dimension has in the trajectory of another dimension. As an example, it would be possible to infer that the corn trajectory, i.e., its path over the years, is similar, or closer, to the CBOT trajectory in dimension *B*. A possible way to verify this is to compute the similarity between trajectories using, e.g., a distance metric.

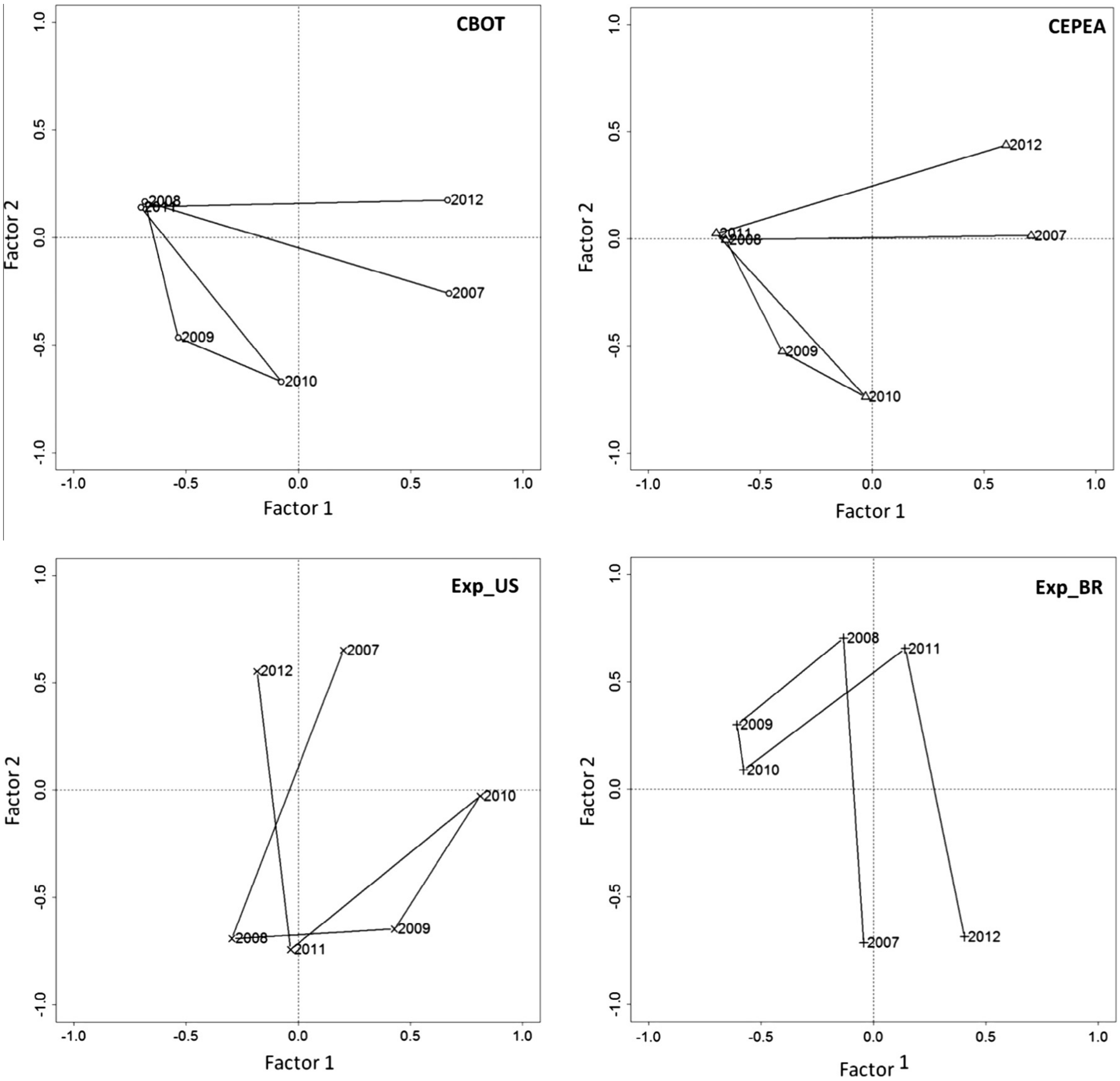


Fig. 6. Four trajectories of dimension *B* – market.

5. Related work

The related papers follow two lines guiding the techniques used on this paper. The first line deals with multidimensional data analysis, in which the main techniques are PCA, PARAFAC and Tucker decomposition model. The second line applies plotting techniques to present multiple results obtained in a bi-dimensional space using maps or trajectories.

The approach for Tucker decomposition and PARAFAC was provided by Frentzos et al. (2009) where it was highlighted the models of PARAFAC and Tucker3 as efficient alternatives for more comprehensive analysis of multidimensional data, when compared to PCA techniques. This is related to the fact that tensor decomposition models, such as PARAFAC and Tucker3, do not collapse the original data dimensions into matrices and are able to explicitly take into account the two and three-way interactions established among these dimensions in the modeling and data decomposition steps. As a result, these techniques do not entail loss of information as happens when using the PCA over the matricized form of three-order tensors. The experiments of Gere et al. (2014) analyzed the decision preferences to buy sweet corns divided by several clusters (organized by people's age). The results were presented by intuitive maps, using tucker3 model by dimension, and compared with the PARAFAC model. However, the paper did not present a solution to deal with time series, or analyzed the time dimension, even though the evolution of such information may reveal important patterns. On the other hand, an approach discussed in Oliveira and Gama (2013), presents a methodology to track the evolution of dynamic social network, using advanced concepts of spatio-temporal trajectories.

Several aspects were applied from these related papers. However, in this paper we covered some innovative points, as for example, real agro economics data, and several data-cubes and in special we used the tucker3 variability to create the trajectory over time series. The proposed methodology presented in this paper allow us to analyze the evolution of multi-cubes (in this case, cubes by year) of time series and projecting them in a representative bi-dimensional space.

6. Conclusion

In this paper, we proposed to use two complementary techniques – Tucker decomposition and analysis of spatio-temporal trajectories – as means of gaining insight about important events related to grain prices time series. The introduced methodology is suitable for the analysis of large amounts of data that arise from agricultural activities. The application of this methodology to a real-world database allowed us to draw some interesting conclusions that, otherwise, would be hard to find. Through the detection of irregularities in data, these techniques can help the experts to focus and concentrate efforts in specific products or markets.

The results using trajectories allowed us to analyze a large amount of multidimensional data, such as the dataset we use, which comprises information regarding 4 products, 2 countries' markets and six years of data. In this context, the trajectories provided an efficient visualization of the evolution of those entities year by year.

Some important aspects were detected by resorting to trajectories: first, both products, soybean and corn prices, had opposite trajectories, which allowed us to infer that these two products will compete for fields in the next crops. On the market analysis, the trajectory of Chicago Stock Market spread the behavior of the prices in the Brazilian domestic market, and both trajectories were similar over the years.

As future work, we intend to improve the analysis with economic aspects and decision making that can be used for analysts and producers. Another issue we want to address is how to compare the similarities of the trajectories over the dimensions. It means to create scenarios about the impact that the trajectory of one dimension has in another trajectory from the next dimension. Further, we intend to use distance metrics, as Euclidean distance or dynamic time warping, to assess the similarity of the trajectories.

Acknowledgments

Center of Advanced Studies on Applied Economics (CEPEA) – University of São Paulo/Brazil.

Laboratory of Artificial Intelligence and Decision Support (LIAAD) – University of Porto/Portugal.

Agricultural Automation Laboratory (LAA) – University of São Paulo/Brazil.

Márcia D.B. Oliveira gratefully acknowledges funding from FCT through a PhD Grant (SFRH/BD/81339/2011).

References

ALICEWEB. The System of Analysis of Foreign Trade Information. Available at: <http://aliceweb.mdic.gov.br/>.

Aruga, Kentaka, 2014. An intervention analysis on the Tokyo Grain Exchange non-genetically modified and conventional soybean futures markets. Cogent Econ. Fin. 2 (1), 918854.

Correa, F.E., 2009. Livestock and agriculture market representation through data warehouse model. Dissertation (Master) – Escola Politécnica, Universidade de São Paulo, São Paulo, 70p.

Datar, M., Gionis, A., Indyk, P., Motwani, R., 2002. Maintaining stream statistics over sliding windows. SIAM J. Comput. 31 (6), 1794–1813.

Frentzos, E., Theodoridis, Y., Papadopouloa, A.N., 2009. Spatio-temporal trajectories. Encyclopedia of Database Systems. Springer, US, pp. 2742–2746.

Gere, Attila et al., 2014. Applying parallel factor analysis and Tucker-3 methods on sensory and instrumental data to establish preference maps: case study on sweet corn varieties. J. Sci. Food Agric. 94 (15), 3213–3225.

IBGE. Brazil Institute of Geography and Statistic. Available at: <http://www.ibge.gov.br>.

Kazmier, L., 2004. Schaum's Outline of Theory and Problems of Business Statistics. McGraw-Hill.

Kiers, H.A., Mechelen, I.V., 2001. Three-way component analysis: principles and illustrative application. Psychol. Methods 6 (1), 84.

King, Robert P. et al., 2010. Agribusiness economics and management. Am. J. Agric. Econ. 92 (2), 554–570.

Kolda, Tamara G., Bader, Brett W., 2009. Tensor decompositions and applications. SIAM Rev. 51 (3), 455–500.

Oliveira, Márcia, Gama, João, 2012. A framework to monitor clusters evolution applied to economy and finance problems. Intell. Data Anal. 16 (1), 93–111.

Oliveira, Márcia, Gama, João, 2013. Visualization of evolving social networks using actor-level and community-level trajectories. Expert Syst. 30 (4), 306–319.

Patel, Dhaval, 2005. On discovery of spatiotemporal influence-based moving clusters. ACM Trans. Intell. Syst. Technol. (TIST) 6 (1), 4.

Pebesma, Edzer, 2012. Spacetime: spatio-temporal data in r. J. Stat. Softw. 51 (7), 1–30.

Plant, Richard E., 2012. Spatial Data Analysis in Ecology and Agriculture Using R. CRC Press.

Rosa, Isabela Ferreira, Bergamin, Leonardo, Makiya, Ieda Kanashiro, 2014. Integration of the soybean production chain and biodiesel: an international parallel to the Brazilian biofuel. Int. J. Innov. Sustain. Dev. 8 (1), 27–36.

Skillicorn, D., 2007. Understanding Complex Datasets: Data Mining with Matrix Decompositions. Chapman and Hall/CRC, New York.

Tucker, Ledyard R., 1966. Some mathematical notes on three-mode factor analysis. Psychometrika 31 (3), 279–311.

Xiao, Ying et al., 2014. Modeling the spatial distribution of crop sequences at a large regional scale using land-cover survey data: a case from France. Comput. Electron. Agric. 102, 51–63.