Cristian KEVORCHIAN, Camelia GAVRILESCU,
Gheorghe HURDUZEU

Institute of Agricultural Economics, Romanian Academy, Bucharest
ck@fmi.unibuc.ro

AN APPROACH BASED ON BIG DATA AND MACHINE LEARNING FOR OPTIMIZING THE MANAGEMENT OF AGRICULTURAL PRODUCTION RISKS

ABSTRACT

Colin Hay and Tony Payne developed the "great uncertainty" concept in order to synthesize certain aspects in relation to the dynamics of economic processes in the current period. They highlight three key elements that mark the structural changes of the moment and that generate incertitude: financial crisis, global economic power change and environmental threats, which bring about significant structural changes in the economy.

At the same time, a technological paradigm change can be noticed at global level, characterized by an unprecedented growth of the "informational power" (computing power, storage power, high analysis capacity) oriented towards the creation of "anti-fragile" economic and administrative structures capable to develop rigorous analyses to outline the tendencies in this area dominated by uncertainty.

The dual solutions in relation to hedging the agricultural commodities supplied by the OTC markets through the weather derived products and those supplied by the insurance companies based on weather indices are well known. We are facing a situation in which the hedging solutions based on OTC market products suffer from a deficit of image following the 2007/2008 crisis, and the solutions provided by the insurers raise the final price by up to 10% depending on the insurance market of each country, playing an important role on the reinsurance market as well.

A unification variant of these markets is presented, providing for operational and production hedging for the farmers acting in a global agricultural industry worth 3 trillion USD, for which 150 billion observations is processed in connection with the soil types, summing up 200 TB of data each month. Weather events are simulated and risk is calculated on 4×4 km^2 areas. These data are combined with business data to support farmers' management decisions.

All these functionalities are supported by a package of technologies based on: NoSQL databases, including HBase, Hive, MySQL, Excel, by software service suppliers in cloud computing context such as Amazon S3, as well as applications delivered under SaaS regime.

A recent report of the Global Institute Mc Kinsey refers to machine learning as an innovation stimulating technology. Being classified as a "silent technology", this can be applied in modeling the weather phenomena to obtain the hedging level necessary to the production context. A composite index such as the Selyaninov index used in a technological context of machine learning type can bring about certain advantages in production risk evaluation. The paper investigates the capacity of certain technologies from the "Big Data" category to add value in the development of certain risk markets in order to obtain a proper hedging.

Key words: agriculture, weather risk, technologies, big data, machine learning.

JEL Classification: Q10.

1. INTRODUCTION

Agriculture has always been considered a “highly risky game”, but the increased number of “extreme weather events” made the planting and harvesting of crops induce unprecedented risks in the farming business. There are areas in the world where the absence of the investment securization possibility in crop establishment, maintenance and harvesting are making a lot of victims. In India, for instance, the delayed monsoon may result in significant harvest losses, which implicitly lead to the impossibility to cover the investment. The result is dramatic, 15000 suicides each year due to the impossibility to compensate the costs of partial or total harvest losses.

It is wrong to consider that the risk coverage system is based on the weather forecast; practically the occurrence probability of certain extreme weather events unfavourable to crops is determined at plot level, as well as the possible impact upon the agricultural farm.

The suppliers of online insurance services for the diminution of risk exposure of agricultural businesses use weather models, climate trends, as well as soil characteristics on the basis of which the agro-pedo-climate data are analyzed at plot level. These analyses are subsequently used to provide farmers with personalized insurance policies against damages or production diminution due to unfavourable weather conditions.

The extremely simple access and practically without involving any bureaucracy to these risk coverage tools determine us to say that these will make it possible to unify into a single electronic market all the range of weather insurance products dedicated to agriculture from the different markets in which they are being operated more or less successfully at present. It is sufficient to underline that farmers do not need to prove their losses, and the payments are triggered off automatically through the electronic payments systems agreed by the parties. For example, the premium practiced by TWI[1] (**T**otal **W**eather **I**nsurance) is 74 USD/ha, and it can amount to 745 USD/ha as compensation for unfavourable weather conditions.

The range of technologies developed round the Hadoop platform can be used to support farmers in the sense of farm protection and operational optimization through software systems of great computational power, dedicated to production risk limitation. The technological platform can combine local weather data, data resulting from agronomic monitoring, as well as data resulting from the application of highly accurate simulation models in a solution that helps farmers to consolidate their profits; furthermore, it leads to better information on decision making and financing.

The financial product dedicated to risk coverage is provided to farmers online and directly impacts their profit. The supplier of services for risk coverage in agriculture should be authorized according to current legislation.

[1] The product of Climate Corporation – Monsanto.

In the face of the increasingly volatile weather conditions, the global business in agriculture that amounts to about 3 trilion USD [13] can be supported by adaptive technological solutions that target profit stabilization and improvement, and, last but not least, the food security of certain large zones in the world.

The financial products of the futures contract type operated by Chicago Mercantile Exchange are well-known, based on temperature and precipitations under the form of snow, yet their volume has been under significant decline, from the record figure of 36 billion dollars in 2005. Within this supply range we can also find the project designed by the authors of the present study based on Selyaninov indices, which besides the operational limitations in data supply, is confronted with the problems of the markets on which hedging is practiced. The administration expenditures for the financial products or of products coming from the insurance market are inducing administration costs that are much too high to be attractive in the countries characterized as "low income countries", and the online insurance solutions can be a suitable solution for this category of countries.

The presented solutions are quite successfully experimented in the USA, and if the combination agriculture-technology proves to be successful, it would be extended globally. Australia, Canada and Brazil may be the next countries that could implement the hedging system based on intelligent analytics provided by the Big Data systems.

2. MATERIAL AND METHODS

Cloud Computing creates the premise of IT distribution as service, this being essentially a metaphor for a family of convergent technologies over Internet through which computational and informational resources are shared, which in their turn are delivered to users as services.

Cloud computing development is based on refining the virtualization technologies and interoperability standards, on growth of universal bandwidth and also at a higher connectivity level facilitated by the shift to Web 2.0. Cloud computing is the continuation of certain technological trends such as grid computing, clustering or server virtualization (Hay & Payne, 2014).

The architectures based on services that target the software infrastructure or systems provided through cloud computing are characterized by:

- flexibility;
- easy recovery in case of disasters;
- automatically updated software;
- no capital costs;
- centralized management;
- security.

Hence, Cloud Computing is a model of computing architecture that enables the access, on demand, through a computer network, to a common pool of IT resources. These resources can be fast and easily identified through the interaction with a provider of resources and services. Among the essential characteristics, we can identify the following: service provided on demand, access through network, pooling the resources, elasticity, control and optimization of resources. The models of services that we can identify in cloud computing are the following:

a) **Software as a Service (SaaS)** – the informatic applications and the related data are stored in a data center and are provided to users on demand, via internet (with a specialized navigator). This service is used for collaborative applications, mobiles, etc., less for real-time applications.

b) **Infrastructure as a Service (IaaS)** – a set of hardware components (servers, storage media, networks, etc.) together with certain software components (operating systems, virtualization, clusterization, etc.) that are offered to users. Infrastructure as a Service offers a family of functionalities in the sphere of basic infrastructure, such as elastic computing or storage services, the clients having the possibility to run any working volume on a cloud infrastructure.

c) **Platform as a Service (PaaS)** – the media for the development and implementation of informatic applications are provided to developers. Platform as a Service offers a complete set of cloud services that make it possible for developers to build complex applications and for customers to use them. The services specifically include:

– Database services for data management and for building Oracle, SQL-Server, MySQL etc. database applications;

– Java, C#, Pyton services, to develop, implement and manage applications in enterprise context;

– Services that facilitate a complete environment of application development such as those for Machine Learning;

– Service for documents management;

– Business Intelligence Service to facilitate the business intelligence capacities in cloud;

– Mobile service that simplifies the access to the mobile applications in cloud.

d) **Data as a Service (DaaS)** – Data quality can be assured on a centralized basis making them available for applications or users. It is an important opportunity to separate the costs of data from the costs of a software or of a specific platform.

As implementation methods we mention:

1) **Private cloud –** the infrastructure is available only inside an organization that includes several users. It can be the case of a network providing insurance. The infrastructure can be owned, configured and used by the respective organization or by third parties, or a combination of these.

2) **Community cloud** – the infrastructure is shared between several entities with common concerns. It can be the example of emergency services – police, fire department, ambulance.

3) **Public cloud** – the infrastructure is for public use for academic or governmental purposes. It presupposes the existence of a third party that physically provides the cloud infrastructure.

4) **Hybrid cloud** – the infrastructure is a combination of private, community and public services that preserves their unitary character, but is united through a technology providing for the portability of information and utilized software applications.

The payment of in cloud services is made according to the *"pay as you go"* principle, and thus the IT services do not involve capital costs, but only operational costs.

In close connection to the cloud computing concept we have the Big Data concept that implies the storage of large volumes of data that are transmitted through specialized protocols that can be processed in real time.

The complex of technologies identified under the generic name Big Data references very large volumes of unstructured data with a very high growth rate. For instance, the digital universe in 2013 totalled about 2.8 Zb[2] and according to International Data Corporation (IDC) it will total 40Zb in 2020. The elements of the digital universe include historical web data, historical data provided by the social networks, data resulting from accessing the banking, email, video clips, pictures services, or can be generated by different categories of IoT (Internet of Things) equipment.

Practically many data from this category continued to be produced in time, but were insufficiently used to obtain valuable information. About 80% of the unstructured data at company level are insufficiently used. An increasing number of Internet users are posting content on the website, so that the volume of data generated increases on exponential basis. Not only the large data volume but also their diversity calls for a fast change of paradigm. In this context, the Big Data concept emerged (Kevorchian *et al.,* 2013).

At present, an increasing number of companies are looking for solutions to most efficiently analyze the data from agriculture in order to be able to make important decisions on their basis and to better understand the phenomena that make agriculture be one of the most sensitive and most volatile investments at global level. A new culture of agricultural data collection and processing is needed, for a more accurate management of processes in agriculture.

Although the volume is the most noticeable characteristic of data coming from and to agriculture, this is not the only problem in making a comprehensive analysis of phenomena in agriculture. There are other two characteristics that mark the data with which the Big Data systems for agriculture are operating: their modification speed (velocity) and the variety of data types sintetized as 3V. These are characteristics that impose a change in the philosophy of data processing in

[2] 1 Zb = 1024^4 Gb.

agriculture. Global companies such as Monsanto and John Deer have understood the importance of such an approach and are making research into this phenomenon summing up several billion USD.

The data generated in the digital universe are very many, and they increase each day, practically hundreds of terabytes being generated every day. McKinsey Institute estimates that the data volume increases up to 40% per year and it is going to increase 44 times in the next 10 years.

All these data circulate with an increasing speed in order to meet the requirements of the current management processes, being generated in real time (for instance the data coming from the social media represent a great flow of opinions and valuable ideas in the management of relations with customers).

The variety of the data delivery format tends to be very rigorously defined and to change as methodology. The data delivery format in a non-traditional manner presents an increasing adoption rate. Depending on the new services made available, new data types emerge in order to capture the informational universe under permanent expansion.

In the face of such a large volume of data, both structured and unstructured, it is difficult to obtain scientifically consistent information out of them. The key information can be very easily lost among the non-significant data. The challenge consists in the identification of truly valuable data, together with the process of their extraction and transformation so as to be able to substantiate a pertinent analysis.

The data can be considered as part of a Big Data approach when the volume, the speed and variety exceed the processing, storage and analysis capacity of traditional systems. This can be solved up on the short term, but when the problem appears time after time, a Big Data solution is needed (Kevorchian *et al.,* 2014).

We can identify four components in a Big Data solution, namely: data collection, data organization, data analysis and the decisions stemming from the implementation of the first three components. The unprocessed data are collected and stored in relational databases, regular data files, files distributed over the Hadoop cluster nodes or in NoSQL databases.

The technologies used are based on Hadoop Distributed File System, HData and NoSQL database organized by columns and database management system as SQL Server, Oracle Database, MySQL, deleiverd as service etc. These are subseq uently filtered and organized through the programming paradigm over HDFS MapReduce, in order to remove the inconsistencies. They can be introduced into relational databases or into a data warehouse in order to be analyzed for business decision making. The technologies used are Hadoop Software Framework, Oracle Loader for Hadoop, Oracle SQL Connector for HDFS and Oracle Data Integrator for Hadoop. For the decision part, the utilized technologies are Oracle Business Intelligence Tools and Oracle Endeca. In the Microsoft AZURE context, Hadoop cluster is consumed as HDInsight data service through a virtual machine provisioned in Azure Cloud OS.

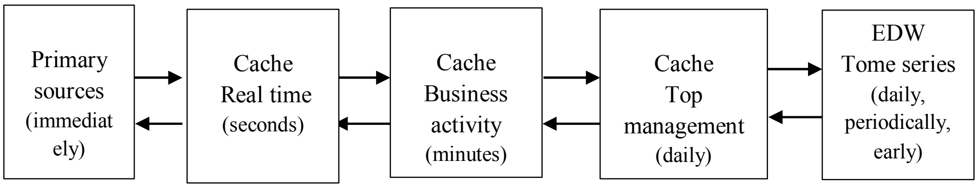
3. ARCHITECTURE OF THE UNIFIED SYSTEM INTO A GLOBAL ONLINE INSURANCE MARKET

A new generation of companies emerged to supply advanced processing of large volumes of data collected in real time with specific information technology tools such as Cloud Computing and Big Data, which cannot be processed with the current technologies of data collection, analysis and presentation. In this category we can mention the suppliers of online services that provide insurance products against weather risks in agriculture.

This hedging form represents a new form of financial product dedicated to farmers and it is included in the class of weather derivatives (which was developed in mid' 1990s, having a large audience around 2005, to experience a great decline in the year 2008), the purpose of which is the insurance against non-extreme weather events. The payment is made regardless of the loss size, if certain conditions are met. Practically, the product can be classified as belonging to the class of financial derivatives rather than to the class of insurance products.

This type of financial product is closely linked to the existence of the global network accessible to everybody (Internet), as well as to the computational power price decrease and increase of farmers' digital education level.

The complexity of data processing at the level of such a system implies finding certain solutions that combine processing variants characterized by small latency (real-time processing) with complex computational processes in the EDW(Enterprise Data Warehouse) zone with great latency but complex transactions as seen in the figure below:


Primary sources (immediat ely)
Cache Real time (seconds)
Cache Business activity (minutes)
Cache Top management (daily)
EDW Tome series (daily, periodically, early)

Source: Kimball, 2012, p. 4.

Figure 1. BIG DATA hybrid solutions.

In Khandany *et al.* (2010), the insurance index for weather risk is strongly correlated with agricultural production losses that implicitly represent a proxy for economic losses. The index utilization is considered by the authors as opportune for the small-income countries, but it induces a boomerang effect in the sense that the effect of climate changes increases the insurance price due to the increase of climate risk. In Khandany *et al.* (2010) we find the index equation:

$$price=risk\_cost + risk\_cost\_loading + administration\_cost + capital\_\ access\_\ cost$$

where the *risk_cost* is modelled *with the linear compensation function* for the weather index associated insurance contract.

The weather data are used for the calculation of a probability distribution associated to the basic weather variable. If the contract protects the insufficient achievement of the basic level of weather variables cumulated as effect, the function takes the following form:

$$compensation = f(i|x, i^*, \lambda) = \begin{cases} 0 & i > \lambda \\ \dfrac{i^* - i}{i^*(1-\lambda)} & \lambda i^* < i \leq i^* \\ 1 & i \leq \lambda i^* \end{cases}$$

where *i* is the achievable value of the basic weather variable, *x* is the sum insured, *i** is the triggerer, and λ Є (0,1) is a threshold value.

The risk cost loading means that the insurer expenditures at an erroneous estimation of weather risk and which is materialized in the reinsurance contract. The administrative costs are expenses incurred with product administration, while the *capital access costs* represent the capital accessed on the international reinsurance market so as to make the product functional.

In Kevorchian *et al.* (2013), a hedging based on the Selyaninov index value is proposed:

$$\mathrm{SHR}_{\mathrm{wheat}} = \frac{\sum_{\mathrm{April\,15\,-\,June\ 30}} \mathrm{Daily\ rainfall}}{0{,}1 * \sum_{\mathrm{April\,15\,-\,Jume\,30}} \mathrm{Daily\ average\ temperature}}$$

where 1 ≤ SHR ≤ 1.4

When the index exceeds the value 1.6, production is lower due to excessive moisture, and when it is smaller than 0.6, production decreases due to the excessive drought.

According to a scenario in which a farmer would buy an insurance policy for the protection against damages caused by eventual flooding, which is a high risk event yet with low probability, the production losses caused by the unfavourable weather conditions would not be included. The impact (upon production, for instance) is determined with the formula below:

$$I(SHR) = \begin{cases} \max\{M, (1 - SHR) * \theta\} & SHR \in [0,1) \\ 0 & SHR \in [1,1.4] \\ \max\{M, (SHR - 1.4) * \theta\} & SHR \in (1.4,2) \end{cases}$$

where M is the value of contract, and θ calculated with the formula

$$F_{wheat}(x) = 3.943x^2 - 1.858x^3$$

is 9.9014 as value of index (the step being 0.01).

The whole chain of operations from the sale of product to the eventual compensations is made without any bureaucratic process. The Report of Climate Policy Initiative (Buchner *et al.*, 2014) reveals that the sums allocated for limiting the extreme climate effects in the period 2011-2013 continually decreased, from billion USD 364 billion in 2011, down to USD 359 billion in 2012 and USD 331 billion in 2013, compared to the necessary USD 5 trillion mentioned by the same source.

From the presented facts, it results that cheap products for the limitation of weather risks still represent an objective that is too difficult to reach. The hope resides in the electronic market area, which by the tremendous progress achieved in cloud computing – Big Data can provide weather risk coverage at competitive prices.

We should not neglect the increasing presence of "automated transactions" on the market. The automated transactions represent the next step in the evolution of trading activities. If e-commerce has made the small or big business operate either online or disappear, how could we avoid the shift to automated transaction?

Initially the "Transaction Programs" emerged, and then other levels of transaction process abstractization were used, which implied an advanced algorithmization of transactions. The utilization level of the technology in the transaction processes depends on the human capacity to abstractize the process, and on the basis of this process to generate performant algorithms, which are interconnected and offer a transaction process of a higher quality compared to that practiced on the market. According to Kaustabh Ray's opinion (Equity Research Technologist and System Arhitect) the transactioning opportunity can be analyzed in the following directions:

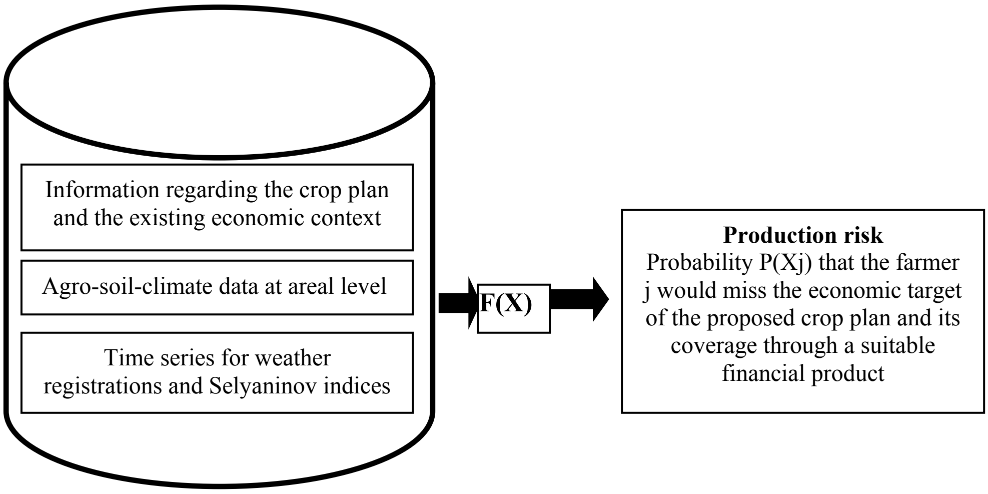
1. Identification of trading opportunities. The result is information sustaining a certain transactional orientation.
2. Orienting a call on the basis of information from 1.
3. If (2) is achieved, then the transaction life cycle will be monitored.

The automation process is based on processing an extremely large volume of historical data on the basis of which rules 1-3 are generated. Through sophisticated machine learning systems, transactioning strategies can be generated characterized by analytical systems that should correspond to market exigencies. The automated transactioning would also result in the removal of bureaucratic processes on the market, which is absolutely necessary.

4. INDEX IMPLEMENTATION ON THE BASIS OF A WEATHER ANALYTICS SYSTEM

The new IT systems based on cloud computing and big data make available software resources that make it possible to analyze large data volumes in order to obtain weather analytics in real time to measure the production risk on the farms in a given area. The weather analytics are computed on the basis of the data collected both from the public and commercial weather stations and from the satellites specialized in such operations of data collection and transmission. The weather data are integrated with soil information as well as with data collected by sensors on the soil temperature and moisture for operations regarding the sizing of the water volume that should be applied through irrigations.

For the financial markets, the analytics calculated in real time are key elements on which the transactioning algorithmics is based. For instance, solutions based on the NoSQL databases such as MongoDB make it possible the loading and fast query of data coming from multiple heterogenous sources. On the basis of a machine-learning algorithm that is based on agro-soil-climate data collected in real time, as well as on the economic context, the farmer's production risk can be calculated.


Information regarding the crop plan and the existing economic context
Agro-soil-climate data at areal level
Time series for weather registrations and Selyaninov indices
F(X)
Production risk
Probability P(Xj) that the farmer j would miss the economic target of the proposed crop plan and its coverage through a suitable financial product

Source: Khandany *et al*., 2010.

Figure 2. Algorithm for the production risk computation.

For any machine-learning model, consistent relations are identified from a logical point of view, between certain characteristics of the entry data and the target variables of the model. Supervised learning is one of the most frequently utilized machine-learning techniques, where the training process is based on entry/ exit associations, where the entries represent attributes with instances associated to certain values of exits. Supervised learning can be considered similar to a regression where the exit can be a continuous or discrete variable. The supervised learning may be seen as similar to a regression, where the exit may be a continuous or discrete variable. The entry is a vector whose components are mapped on exits. The entry is a vector whose components are mapped on exits. The entry/exit pairs are values that define the training process. Through a "brute force" method all the entry/exit pairs are memorized. From the training data “the noise” should be removed, so as not to affect the process of obtaining the result function. An output of our model is a continuous variable with the values belongs to the unit interval that measures the risk exposure of a given farmer, by using a Bayesian model present in most machine-learning systems. The utilization of a metrics family is needed, and depending on the result of testing, adjustments can be made through calibration. In the practice of farm production risk management, a cost/benefit analysis is used, which can be used as a witness in running the modelling process.

The model operates with data taken over from a Hadoop cluster through MapReduce architecture, which allows both the automatic parallelization and a distribution of both tasks and data over the cluster. In his article, Raden (2012), shows that through the second type of platform, specialized libraries in statistical routine and machine learning are supplied, as well as the cluster management possibility. For operational analytics, at the Hadoop cluster level, the storage and integration facilities are used. The database queries use a HiveQL language or are based on third-party software products. For analyses of Business Intelligence type the operational register is changed from the NoSQL typical Hadoop zone to the relational zone, where numerous analytical tools can be identified, as well as connectors to a wide range of data sources.

5. DISCUSSIONS AND CONCLUSIONS

The range of technologies developed around the Hadoop platform can be used for farmers’ support, in the sense of farm operational protection and optimization through software systems of high computational power dedicated to the production risk insurance. The technological platform can combine local weather data, data resulting from agronomic monitoring as well as data resulting from the application of certain highly accurate simulation models (machine learning) into a solution that helps farmers to consolidate their profits; moreover, it can lead to a better information on decision making and financing.

The suppliers of online insurance services that design IT products meant to lower the risk exposure of farming business use weather models, climate trends, as well as soil characteristics on the basis of which they analyze agro-soil-climatic data at plot level.

These analyses are subsequently used to offer personalized insurance policies to farmers against damages or production losses caused by unfavourable weather conditions at plot level.

The extremely simple access and practically null bureaucracy of these risk coverage tools enable us to declare that the whole range of weather insurance products dedicated to agriculture (coming from different markets on which they have been operated more or less successfully so far) will be able to unify on an electronic market.

REFERENCES

1. Buchner B., Stadelmann M., Wilkinson J., Mazza F., Rosenberg A., Abramskiehn D. (2014), *The Global Landscape of Climate Finance*, Climate Policy Initiative, Venice, Italy (http://climate policyinitiative.org/wp-content/uploads/2014/11/The-Global-Landscape-of-Climate-Finance-2014.pdf).
2. Collier B., Skees J., Barnett B. (2009), *Weather Index Insurance and Climate Change: Opportunities and Challenges in Lower Income Countries*. The Geneva Papers – issues and practice, vol. 34, pp. 401-424.
3. Hay C., Payne T. (2013). *The Great Uncertainty.* Sheffield Political Economy Research Institute, SPERI Paper no. 5 (http://speri.dept.shef.ac.uk/wp-content/uploads/2013/01/SPERI-Paper-No.5-The-Great-Uncertainty-389KB.pdf).
4. Bathes J. (2014), *Climate Risk, Big Data and the Weather Market.* Sheffield Political Economy Research Institute, SPERI Paper no. 13 (http://speri.dept.shef.ac.uk/wp-content/uploads/2014/05/SPERI-Paper-No.-13.pdf).
5. Kevorchian C., Gavrilescu C., Hurduzeu G. (2013), *Qualitative Risk Coverage In Agriculture Through Derivative Financial Instruments Based On Selyaninov Indices*. Financial Studies, vol. 17(3).
6. Kevorchian C., Gavrilescu C., Hurduzeu, G. (2014), *The Architecture of Informatics Systems for Farm Management – a Cloud Computing and Big Data Approach.* EAAE 2014 Congress (Ljubljana), http://ageconsearch.umn.edu/handle/182844
7. Khandany A., Aldar K., Lo A. (2010), *Consumer Credit Risk Models via Machine-Learning Algorithms*. Journal of Banking & Finance, vol. 34, p. 2767-2787.
8. Kimball R. (2012), *Newly emerging best practices for Big Data*, (http://www.kimballgroup. com/wp-content/uploads/2012/09/Newly-Emerging-Best-Practices-for-Big-Data1.pdf).
9. Loshin D. (2013), *Big Data Analytics. From strategic planning to enterprise integration with tools, techniques, NoSQL and graph*, Elsevier, Morgan Kaufmann imprint.
10. Gulati M., Fulay A., Datta S. (2013), *Building and Managing a Cloud Using Oracle Enterprise Manager 12c*, Oracle Press.
11. Raden N. (2012), *Big Data Analytics Architecture.* (http://www.thebigdatainsightgroup. com /site/sites/default/files/Teradata's%20-%20Big%20Data%20Architecture%20-%20Putting%20all%20your%20eggs% 20in%20one%20basket.pdf).
12. Plunkett T., Macdonald B., Nelson B., Hornick M., Sun H., Mohiuddin K., Harding D., Mishra G., Stackowiak R., Laker K., Segleau D. (2013), *Oracle Big Data Handbook.* Oracle Press.
13. http://en.wikipedia.org/wiki/The_Climate_Corporation#