Improving Modeling and Forecasting of Fuel Selling Price Using Support Vector Machines: Case Study

Zineb Aman, Haj EL MOUSSAMI, Younes Fakhradine EL BAHI
Mechanics & Integrated Engineering
ENSAM School, Moulay Ismail University
Meknes, Morocco
zineb.aman@gmail.com, elbahiyounesfakhradine@gmail.com, hajelmoussami@yahoo.com

Latifa Ezzine
Modeling, Control Systems and Telecommunications
Faculty of Sciences, Moulay Ismail University
Meknes, Morocco
latifae@yahoo.com

Abstract

The liberalization of the petroleum sector in Morocco has a significant effect for petroleum product distributors. Since the beginning of December 2015, fuel prices are freely determined. This event presents many constraints affecting the balance of the sector plus the competition between its economic players. With the halt of the competitive manufacturing's activity, Morocco's only refinery, distributors must, for their part, build up large stocks. As all fuel products are imported, we will be interested in the evolution by making forecasts of the price of fuels in the Moroccan market. In order to achieve their objectives, the oil companies must rely on precise forecasts. In this context, our paper aims mainly to study the time series of fuel selling price in order to provide precise forecasts to the company respecting the permissible error margin of 3%. To this end, we worked with the SVR function. The predictions made were quite satisfactory with regard to the constraint set by the company (plus or minus 3% as margin of error). The error of the SVR function used is about 2,53%. The error in then minimized compared to our previous method: ARIMA which error is about 2.855%.

Keywords
liberalization of the petroleum sector; Time series; ARIMA; SVR function.

1. Introduction

Crude oil price forecasting has widely been considered as one of the most important but challenging issues in the research fields of data analysis and prediction, due to the interactive uncertain factors driving the crude oil market. On the one hand, similar to other energy commodities, the crude oil is directly associated with various uncertain market factors, e.g., supply and demand, competition between providers, substitution with other energy forms, economic development, population growth and technique development (Yu et al. 2014). On the other hand, as the crude oil is a dominant resource concerning energy security, the crude oil market is super-sensitive to diverse uncertain external factors, such as political instabilities, wars and conflicts (Yu et al. 2017) (Naser et al. 2016).

Recently, oil prices have made the headlines of the financial press on a daily basis in Morocco. Since the end of 2015, fuel prices are now governed by the free play of supply and demand. Now, consumers have the freedom to choose the station they can use. This liberalization is in principle an opportunity for the Moroccan economy because it will encourage all economic actors to rationalize their behavior. However, the analysis of the government's approach reveals several shortcomings that could prevent consumers from taking full advantage of the benefits of this liberalization. Several risks hover around this measure that will have to be curbed. In this sense, stock uncertainty poses a real threat to supply and a risk of rising prices. What is in question here is rather the timing of the liberalization which is not very opportune with the problematic stop of the production of the only refinery of Morocco ensuring alone about half of the 60 days of the country's strategic stock. Especially that the risk of out of stock with the winter period (disruption of supply due to the bad weather) is even greater and must be taken seriously. Because, in case of pressure on stocks, it will encourage speculation of all kinds and prices will flare up. The uncertainty also relates to the evolution of oil prices. Certainly several studies estimate that the price of oil should not exceed 60 dollars in the 2 years to come.

Nevertheless, nothing is less certain because we are never immune to a turnaround of the economy especially in the current sensitive geopolitical context that can make prices start up again. And in the event of a turnaround, it is clear that the government has not planned anything and believes that this status quo will last as long as possible. The challenge here is to prevent the government from questioning liberalization by intervening on prices in the event of a rise in prices.

The case of the suppression of indexing in 2000 following the rise in international prices is still there to testify. Hence, the need to provide support mechanisms (mutual insurance mechanism and price forecast for example), which would amortize the possible rise in prices. And even if prices are low today, note that with an overvalued dollar, prices at the pump will be higher than they should be.

This was the case for the last fortnight of October 2015, when the price decline was counterbalanced by the rise in the dollar. The risk posed by dollar fluctuations is to be taken. Our paper aims to implement a forecast process for fuel prices, we will try to forecast the fuel named super unleaded “Super sans plomb: SSP” prices for the year 2017 based on the history of the four previous years through the SVR function and compare results with those found when applying ARIMA method.

2. Literature review

2.1. Forecasting Price

Oil price forecasting is known to be a challenging task. The large number of variables affecting the price, the non-linear effects and feedback loops, and all the “unknowns” and uncertainties can quickly compound into very different estimates depending on who you ask. Maybe for this reason the outlooks that are presented and published have been dominantly concerned with the price development for a few years ahead. Only occasionally researchers aim to model oil and other energy prices into the more distant future. Given the importance of crude oil as the world’s primary energy source in general it is crucial to have a good understanding of how these prices behave in the long term, without getting blinded by all the details and noise influencing the short term prices (Haugom et al. 2016).

Despite our great oil dependence, there is remarkably little research aiming at understanding the long term real oil price development. The focus of the research within the oil sector is often on reserves, future supply or (peak) production (Greene et al. 2006), (Kjärstad et al. 2009), (Kerr et al. 2011), (Maggio et al. 2009), taxation (Dasgupta et al. 1980), oil price shocks based on properties of supply and demand (Kilian 2008), (Kilian 2009), or (relatively) short term forecasting of oil prices and consumption (Gori et al. 2007). One reason for why so little attention has been given to long term oil price modeling and forecasting may be that it is known to be difficult due to ambiguous or poor information about the true global oil resources (Kjärstad et al. 2009), and the complexity of both the characteristics of the commodity and the market mechanisms in general. One recent study that focuses on understanding the behavior of oil prices is the seminal paper of Hamilton (Hamilton et al. 2008).

The oil price forecasts is the basis of all supply chain planning. The pull processes in the supply chain are performed in response to customer demand, whereas all push processes are performed in anticipation of customer demand (Chopra et al. 2006). A company must understand such factors before it can select an appropriate forecasting methodology because it may have difficult to decide which method is most appropriate for forecasting. Forecasting methods are classified according to the following four types: Qualitative, Time series, Causal and Simulation (Chopra et al. 2006).

A time series consists of observations arranged in chronological order (Bozarth et al. 2016). Time series forecasting models use mathematical techniques that are based on historical data to forecast oil price. It is based on the assumption that the future is an extension of the past; thus, historical data can be used to predict future oil price (Wisner et al. 2011).

By time series analysis, the forecasting accuracies depend on the characteristics of time series of demand. If the transition curves are stable and periodical the high forecasting accuracies will be expected, whereas high accuracies cannot be expected if the curves show highly irregular patterns (Matsumoto et al. 2015).

However, in recent years, Support Vector Regression has been successfully used in forecasting problems. Support Vector Machines (SVMs) have gained importance in forecasting problems related to the environment (Ornella et al. 2010), (Jain et al. 2009). There are two main categories in support vector machines: support vector classification (SVC) and support vector regression (SVR). SVMs use a high-dimensional feature space (Balahura et al. 2014), (Chakraborty et al. 2011). Support Vector Regression (SVR) was specifically developed and comprises appealing algorithms for a wide variety of regression problems (Rajasekaran et al. 2008), (Yang et al. 2009), (Wei et al. 2013), (Zhang et al. 2013).

2.2. Support Vector Regression

SVM has been successfully used for modeling and predicting time series. The SV (Support Vector) algorithm is a nonlinear generalization of the generalized Portrait algorithm developed in Russia in the sixties (Vapnik et al. 1963).VC theory has been developed over the last three decades by Vapnik, Chervonenkis and others (Vapnik et al. 1974), (Vapnik et al. 1982), (Vapnik et al. 1995). This theory characterizes properties of learning machines which enable them to effectively generalize the unseen data. In its present form, the SV machine has been developed at AT & T Bell Laboratories by Vapnik and co-workers (Vapnik et al. 1997). Initial work has focused on OCR (optical character recognition). Within short period, SV classifiers have become competitive with the best available systems for both OCR and object recognition tasks (Schölkopf et al. 1998). Burges (Burges, 1998) published a comprehensive tutorial on SV classifiers. Excellent performances have been obtained in regression and time series prediction applications (Drucker et al. 1997). Statistical Learning Theory has provided a very effective framework for classification and regression tasks involving features. Support Vector Machines (SVM) are directly derived from this framework and they work by solving a constrained quadratic problem where the convex objective function for minimization is given by the combination of a loss function with a regularization term (the norm of the weights). While the regularization term is directly linked, through a theorem, to the VC-dimension of the hypothesis space, and thus fully justified, the loss function is usually (heuristically) chosen on the basis of the task at hand.

Traditional/statistical regression procedures are often stated as the processes deriving a function f(x) that has the least deviation between predicted and experimentally observed responses for all training examples. One of the main characteristics of Support Vector Regression (SVR) is that instead of minimizing the observed training error, SVR attempts to minimize the generalized error bound so as to achieve generalized performance. This generalization error bound is the combination of the training error and a regularization term that controls the complexity of the hypothesis space. Support vector machine (SVM) has been first introduced by Vapnik. There are two main categories for support vector machines: support vector classification (SVC) and support vector regression (SVR).

SVM is a learning system using a high dimensional feature space. It yields prediction functions that are expanded on a subset of support vectors. SVM can generalize complicated gray level structures with only a very few support vectors and thus provides a new mechanism for image compression. A version of a SVM for regression has been proposed in 1997 by Vapnik, Steven Golowich, and Alex Smola (Vapnik et al. 1997). This method is called support vector regression (SVR). The model produced by support vector classification only depends on a subset of the training data, because the cost function for building the model does not care about training points that lie beyond the margin. Analogously, the model produced by SVR only depends on a subset of the training data, because the cost function for building the model ignores any training data that is close (within a threshold ε) to the model prediction. Support Vector Regression (SVR) is the most common application form of SVMs. An overview of the basic ideas underlying support vector (SV) machines for regression and function estimation has been given in (Smola et al.1998).

Suppose we are given training data {(x1, y1),...,(x, y)} ⊂ X × R, where X denotes the space of the input patterns (e.g. X = Rd). These might be, for instance, exchange rates for some currency measured at subsequent days together with corresponding econometric indicators. In ε-SV regression (Vapnik 1995), our goal is to find a function f (x) that has at most ε deviation from the actually obtained targets yi for all the training data, and at the same time is as flat as possible. In other words, we do not care about errors as long as they are less than ε, but will not accept any deviation larger than this. This may be important if you want to be sure not to lose more than ε money when dealing with exchange rates, for instance. For pedagogical reasons, we begin by describing the case of linear functions f , taking the form

$$f(x) = \langle w, x\rangle + b \text{ with } w \in X,\ b \in R \tag{1}$$

where < · , · > denotes the dot product in X. Flatness in the case of (1) means that one seeks a small w. One way to ensure this is to minimize the norm,[3] i.e. $||w||^2 = \langle w,w\rangle$. We can write this problem as a convex optimization problem:

$$\begin{aligned} &\text{minimize} && \tfrac{1}{2}||w||^2 \\ &\text{subject to} && \begin{cases} y_i - \langle w, x_i\rangle - b \le \varepsilon \\ \langle w, x_i\rangle + b - y_i \le \varepsilon \end{cases} \end{aligned} \tag{2}$$

The tacit assumption in (2) was that such a function f actually exists that approximates all pairs (x_i, y_i) with ε precision, or in other words, that the convex optimization problem is feasible. Sometimes, however, this may not be the case, or we also may want to allow for some errors. Analogously to the “soft margin” loss function (Bennett and Mangasarian 1992) which was used in SV machines by Cortes and Vapnik (1995), one can introduce slack variables ξ_i, ξ_i^* to cope with otherwise infeasible constraints of the optimization problem (2). Hence we arrive at the formulation stated in Vapnik (1995).

$$\begin{aligned} &\text{minimize} && \tfrac{1}{2}||w||^2 + C * \sum_{i=1}^{l}(\xi i + \xi i *) \\ &\text{subject to} && \begin{cases} y_i - \langle w, x_i\rangle - b \le \varepsilon + \xi_i \\ \langle w, x_i\rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \end{aligned} \tag{3}$$

The constant C > 0 determines the trade-off between the flatness of f and the amount up to which deviations larger than ε are tolerated.

Figure 1. depicts the situation graphically.

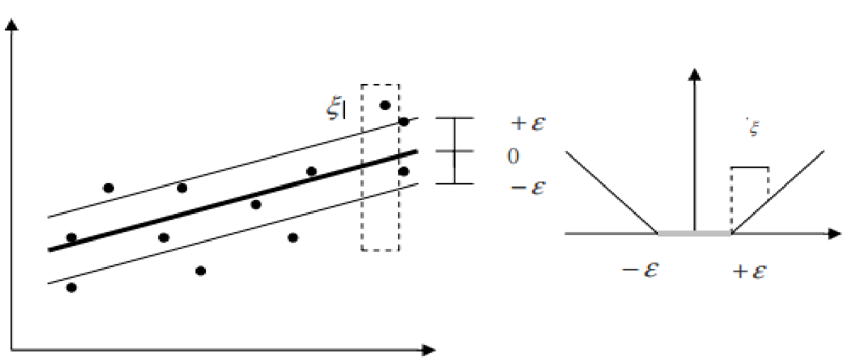

Figure 1. The soft margin loss setting corresponds to a linear SV machine

3. Case Study

In this section, we will model the real data of the price of fuel named « SSP » in order to make predictions that are important to determine future selling prices using SVR (Support Vector Regression) function. This study examines the effectiveness of fuel selling price forecasting to come up with conclusions in terms of the superiority in forecasting performance compared to ARIMA model.

The model shown on Figure 2 is based on the price of the fuel "SSP" in a Petroleum manufacturing from January 2012 until December 2016.

Figure 2. Selling Price of "SSP"

In our previous work (EL BAHI et al. 2018), we conducted a forecasting study of selling price of SSP using ARIMA model. The ARIMA model (1,1,1) is the one that provides accurate forecasts. After having obtained the coefficients, the equation of the model retained is as follows:

$$\boldsymbol{y_t = y_{t-1} - 0.928(y_{t-1} - y_{t-2}) + 0.873\varepsilon_{t-1} + \varepsilon_t} \tag{4}$$

Figure 3 presents results of ARIMA model (1,1,1). Table 1 lists the forecasts obtained for the first quarter of 2017.

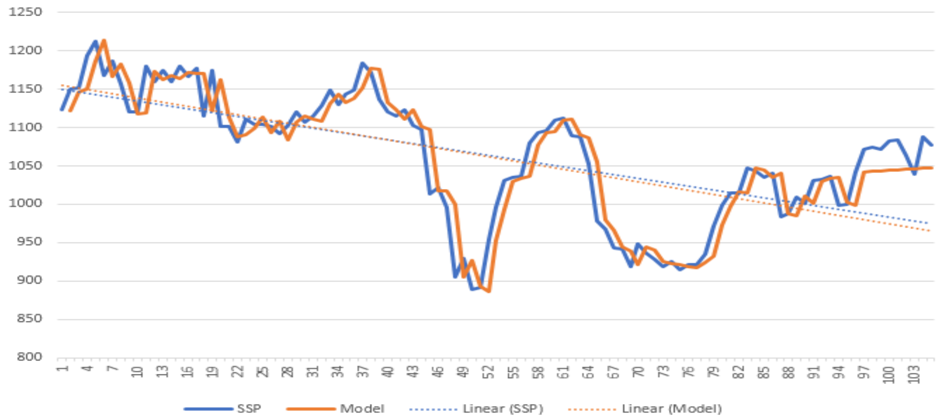

Figure 3. Results of the ARIMA model (1,1,1)

Table 1. Forecast Results for the ARIMA Model (1,1,1)

Fortnight	Real Price	Model	% Error
1Q January	1072	1042,49	-2,752798507
2Q January	1074	1043,05	-2,881750466
1Q February	1072	1043,59	-2,650186567
2Q February	1082	1044,21	-3,492606285
1Q March	1084	1044,81	-3,615313653
2Q March	1064	1045,48	-1,740601504

According to the graph on Figure 3, and the table 1, we noticed that the real model is very close to the one developed at the base of forecasts made using the ARIMA (1,1,1) model. The average error is about 2.855%

The aim of our actual work is to develop a relation between experimental data gathered from authentic sources to estimate fuel selling price. We attempt to apply Support Vector Machines which is based on machine learning approaches because of the complexity of relations between input parameter and the output.

We have prepared our database, and subsequently developed the program in Python language which will be compiled on Spyder software.

We imported our data set which is real prices of our fuel studied, created and indexed the placement for database values. Then, we standardized the data to fit the learning process that will be done using SVR function. In fact, we divided our database in part learning and another for the test. We tried two main distributions: a) 60% of our database used in learning phase and 40% used in test phase, b) 80% of our database used in learning phase and 20% used in test phase. We retained the second distributions based on results obtained after program compilation. After that, we learned (X-train) and (Y-train) and run the test to calculate finally the average of the errors and obtained predicted values in the test phase that are grouped in Figure 4.

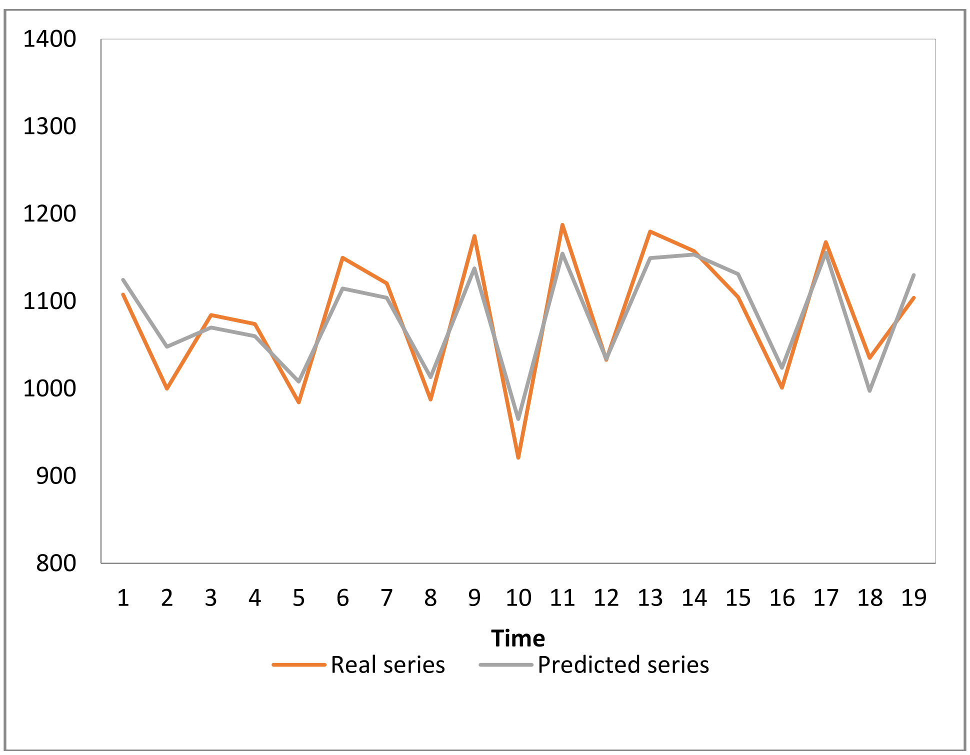

Figure 4. Results of the SVR function

The average error equals 26.882361 that presents 2.53%. The error graph is presented in Figure 5.

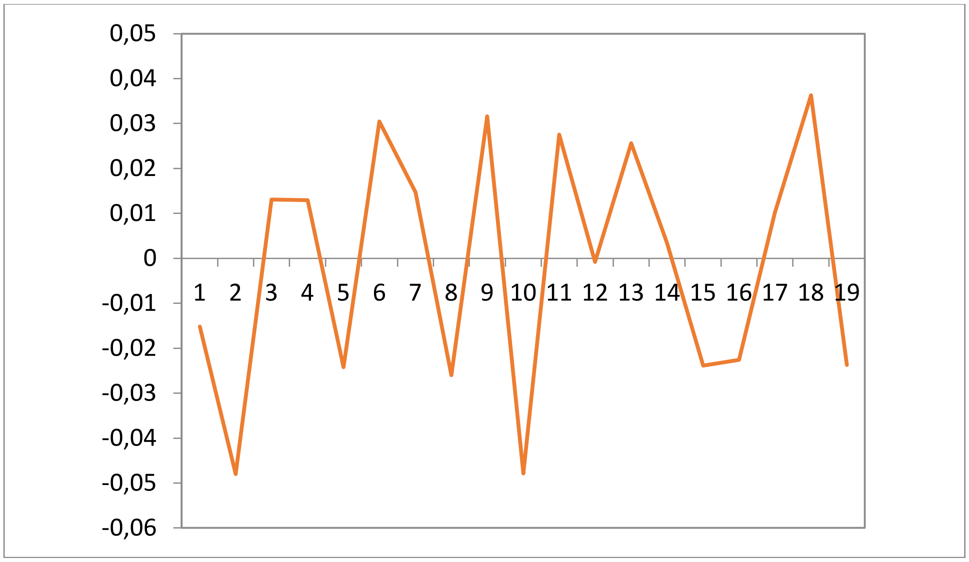

Figure 5. Error graph

We can clearly see that the model chosen can be used for modeling and forecasting the future sells in this petroleum manufacturing since the error found (2.53%) respects the permissible error margin fixed by the company at 3%. Moreover, SVR function is a useful tool which guarantees a good accuracy and minimized the error compared to ARIMA model.

Conclusion

In this paper, we studied the selling prices of the SSP through the SVR function. Our aim is to ameliorate the modeling and forecasting of fuel selling price. To this end, we have prepared our database, and subsequently developed the program in Python language which will be compiled on Spyder software. The predictions made were quite satisfactory with regard to the constraint set by the company (plus or minus 3% as margin of error). The error of the SVR function is about 2.53%. Consequently, the SVR function proved its strength manifested in the error that was further minimized: 2.53 % instead of 2.855% for ARIMA model.

In our future work, we will use another method in this way that will minimize the error more, take into account unforeseen factors and analyze the change of fuel selling price in up and down.

References

Balahura A. and Turchi, M. Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis, *Comput. Speech Lang*., vol. 28, no. 1, pp. 56–75, Jan. 2014.

Bozarth, C. B., Handfield, R. B. Introduction to Operations and Supply Chain Management, North Carolina State University, 2016.

Burges, C.J.C. A tutorial on support vector machines for pattern recognition, *Knowledge Discovery and Data Mining, 2(2),* 1998.

Chakraborty, S. Bayesian semi-supervised learning with support vector machine, *Statist. Methodol.*, vol. 8, no. 1, pp. 68–82, 2011.

Chopra S., and Meindl, P. Supply Chain Management: Strategy, Planning, and Operation – 28 March 2006.

Dasgupta, P., Heal, G., and Stiglitz, J. E. The taxation of exhaustible resources, 1980.

Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A. and Vapnik, V. Support vector regression machines. In M. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems 9,* pages 155-161, Cambridge, MA, MIT Press, 1997.

El Bahi, Y.F., Ezzine, L., Aman, Z., El Moussami, H. Modeling and Forecasting of Fuel Selling Price Using Time Series Approach: Case Study . *Proceedings of the International Conference on Control, Decision and Information Technologies*, Thessaloniki, Greece, April 10-13, 2018.

Gori, F., Ludovisi, D., and Cerritelli, P. F. Forecast of oil price and consumption in the short term under three scenarios: parabolic, linear and chaotic behavior. *Energy, 32(7),* pp. 1291-1296, 2007.

Greene, D. L., Hopson, J. L., and Li, J. Have we run out of oil yet? Oil peaking analysis from an optimist's perspective, *Energy Policy*, 34(5), pp. 515-531, 2006.

Hamilton, J. D. Understanding crude oil prices, *The Energy Journal, International Association for Energy Economics,* vol. 30(2), pp. 179-206, 2008.

Haugom, E., Mydland, Ø., and Pichler, A. Long term oil prices, *Energy Economics*, 58, pp. 84-94, 2016.

Jain, P., Garibaldi, J. M., and Hirst, J. D. Supervised machine learning algorithms for protein structure classification, *Comput. Biol. Chem.*, vol. 33, no. 3, pp. 216–223, 2009.

Kerr, R. A. Peak oil production may already be here, *Science*, 331(6024), pp. 1510-1511, 2011.

Kilian, L. Exogenous oil supply shocks: how big are they and how much do they matter for the US economy?, *The Review of Economics and Statistics*, 90(2), pp. 216-240, 2008.

Kilian, L. Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil marke, *Am. Econ. Rev. 99 (3),* pp. 1053–1069, 2009.

Kjärstad, J., and Johnsson, F. Resources and future supply of oil, *Energy Policy*, 37(2), pp. 441-464, 2009.

Maggio, G., and Cacciola, G. A variant of the Hubbert curve for world oil production forecasts, *Energy Policy*, 37(11), pp. 4761-4770, 2009.

Matsumoto M., and Ikeda A., Examination of demand forecasting by time series analysis for auto parts remanufacturing, *Journal of Remanufacturing*, 5(1), 1, 2015.

Naser, H. Estimating and forecasting the real prices of crude oil: A data rich model using a dynamic model averaging (DMA) approach, *Energy Economics, Volume 56*, pp: 75-87, May 2016.

Ornella L., and Tapia, E. Supervised machine learning and heterotic classification of maize (Zea mays L.) using molecular marker data, *Comput. Electron. Agriculture*, vol. 74, no. 2, pp. 250–257, 2010.

Rajasekaran, S., Gayathri, S., and Lee, T.-L. Support vector regression methodology for storm surge predictions, *Ocean Eng.*, vol. 35, no. 16, pp. 1578–1587, 2008.

Schölkopf, B., Bartlett, P.L., Smola, A. and Williamson, R. Support vector regression with automatic accuracy control, In L. Niklasson, M. Boden, and T. Ziemke, editors, *Perspectives in Neural Computing: Proceedings of the 8th International Conference on Artificial Neural Networks (ICANN'98),* pages 111-116. Springer-Verlag, 1998.

Smola, A.J. and Schölkopf, B. A Tutorial on Support Vector Regression, NeuroCOLT, Technical Report NC-TR-98-030, Royal Holloway College, University of London, UK, 1998.

Vapnik V., and Lerner, A. Pattern recognition using generalized portrait method, *Automation and Remote Control*, 24, 1963.

Vapnik, V. and Chervonenkis, A. Theory of Pattern Recognition, (in Russian), *Nauka*, 1974.

Vapnik, V. and Chervonenkis, A. A note on one class of perceptrons, *Automation and Remote Control,* 25.

Vapnik, V. Estimation of dependences based on empirical data, *Springer Verlag*, 1982.

Vapnik, V. The Nature of Statistical Learning Theory, *Springer Verlag*, 1995.

Vapnik, V., Golowich, S., and Smola, A. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing, in M. Mozer, Jordan, M., and Petsche, T. (eds.), *Neural Information Processing Systems*, Vol. 9. MIT Press, Cambridge, MA, 1997.

Wei, Z., Tao, T., ZhuoShu, D., and Zio, E. A dynamic particle filtersupport vector regression method for reliability prediction, *Rel. Eng. Syst. Safety*, vol. 119, pp. 109–116, Nov. 2013.

Wisner, J. D., Tan, K. C., and Leong, G. K. Principles of supply chain management: A balanced approach, 2011.

Yang, H., Huang, K., King, I. and Lyu, M. R. Localized support vector regression for time series prediction, *Neurocomputing,* vol. 72, nos. 10–12, pp. 2659–2669, 2009.

Yu, L., Xu, H., and Tang, L. LSSVR ensemble learning with uncertain parameters for crude oil price forecasting, Applied Soft Computing, Volume 56, pp 692-701, July 2017.

Yu, L., Zhao, Y., and Tang, L. A compressed sensing based AI learning paradigm for crude oil price forecasting, *Energy Economics 46*, pp: 236–245, 2014.

Zhang, L., Zhou, W.-D., Chang, P.-C., Yang, J.-W. and Li, F.-Z. Iterated time series prediction with multiple support vector regression models, Neurocomputing, vol. 99, pp. 411–422, Jan. 2013.

Biographies

Zineb AMAN is a Ph.D. Student at ENSAM School of Meknes, Moulay Ismaïl University, Morocco. She is an Industrial Engineer. She received her Engineering Degree in Industrial Engineering in 2015 from the ENSAM School (Ecole Nationale Supérieure d'Arts et Métiers de Meknès), Moulay Ismaïl University, Morocco. Her research interest includes modeling and optimization, supply chain management, performance evaluation, continuous improvement.

Latifa EZZINE is a Professor at Department of Industrial Engineering, ENSAM School (Ecole Nationale Supérieure d'Arts et Métiers de Meknès), Moulay Ismaïl University, Morocco. She obtained her Ph.D. Degree in 2010 from the Faculty of Sciences, Moulay Ismaïl University, Morocco. Her current research focuses on statistics and artificial intelligence, modeling and optimization, control systems and telecommunications, etc.

Younes Fakhradine EL BAHI is a Ph.D. Student at ENSAM School of Meknes, Moulay Ismaïl University, Morocco. His research interest includes optimization, modeling and Simulation of Supply Chain Management.

Haj EL MOUSSAMI is a Professor at Department of Mechanical Engineering, ENSAM School (Ecole Nationale Supérieure d'Arts et Métiers de Meknès), Moulay Ismaïl University, Morocco. He obtained his Ph.D. Degree in 2003 from the ENSAM School (Ecole Nationale Supérieure d'Arts et Métiers de Bordeaux), France. His current research focuses on computer-aided design and manufacturing, manufacturing processes, mechanics and integrated engineering, tribology, etc.