2013 International Conference on Advances in ICT for Emerging Regions (ICTer): 132 - 136

Short-term Forecasting of Electricity Consumption in Maputo

Constantino Sotomane[1,2], Lars Asker[1], Henrik Boström[1], Venâncio Massingue[3]
[1]Stockholm University, Department of Computer and Systems Science, Sweden
[2]Ministry Of Science and Technology, Mozambique
[3]Eduardo Mondlane University, Mozambique
Sotomane@dsv.su.se, Asker@dsv.su.se, henrik.bostrom@dsv.su.se, Venancio.massingue@uem.mz

Abstract. We present a short-term load forecasting model for Maputo. The model is based on the concept of multiple models. A clustering method is combined with expert's knowledge to identify sub-models. The resulting model, which is the combination of several sub-models, is evaluated and compared to the model currently used by the Electricidade de Moçambique E.P (EDM). The results show that the developed model performs better accuracy than the one currently used by EDM. The results obtained by the application of the model when translated into financial figures demonstrate significant economic advantages. The social and environmental implications of the model are also analysed.

Keywords: **Short-term load forecasting, Day-Ahead-Market, robust regression, multiple models, clustering, Mozambique.**

I. Introduction

Load forecasting is important for management and operation in the electricity industry [1-3]. There are three types of load forecasting models: long-term, mid-term and short-term. Long-term load forecasting corresponds to predicting load one year ahead and is used for planning, e.g., the planning of investments in new capacity [4, 5]; mid-term load forecasting corresponds to a period of days to weeks or months and is used to estimate the medium-term load requirements, such as peaks during summer or winter periods [4]; while short-term load forecasting corresponds to the period from hours to days and is important for management and operation of power systems and the electricity market [6].

Short-term load forecasting is the basis for several operations activities such as scheduling the capacity of generation, reliability analysis, security assessment and maintenance plans [1, 4, 6]. With the deregulation of the electricity sector, short-term load forecasting has become more critical due to its importance for planning of electricity transactions in the competitive market [1, 3, 4, 6]. The importance and complexity of short-term load forecasting will continue to increase, due to supply-demand fluctuation, changes in prices of electricity, changes in weather and the high financial penalties resulting from forecasting errors [1, 5].

The structure of the electricity sector has been changing in the last decades all over the world, from monopoly to deregulated and competitive market structures [5]. As part of the restructuration process, the member states of the Southern African Development Community (SADC) created in 1995 the Southern Africa Power Pool (SAPP) *"with primary aim to provide reliable and economical electricity to supply to the consumers of each of the SAPP members, consistent with reasonable utilization of natural resources and the effect on the environment"[7].*

Since 2009, SAPP is operating a competitive energy market for SADC in the form of a Day-Ahead-Market (DAM), where bids for electricity trading are submitted one day before [5, 8]. The Electricidade de Moçambique E.P (EDM), the main supplier of electricity in the country, is member of SAPP and also must send its bids one day in advance in order to participate in the market. To be able to send the bids a day in advance, it is necessary to forecast the demand of the internal consumers and based on that, to infer the amount of electricity for DAM.

EDM forecasts the hourly maximum consumption for the next day and based on that, estimates the amount of electricity for the DAM. This procedure is not optimal since it underestimates the electricity for DAM, resulting in loss of electricity that is not consumed neither sold. The aim of this work is to build a short-term load forecasting model to improve the accuracy of electricity estimated to DAM.

The electricity forecast is done traditionally using statistical methods such as linear or multiple regression, box Jenkins and others, but it has been observed that these techniques are deficient when there are abrupt changes in environmental or sociological variables [1, 2, 9]. In this work, we use the concept of multiple-models where the forecasting model is composed of a set of sub-models. Each sub-model is used to forecasts the consumption's demand of certain type of days, for example the first sub-model can be used to forecast working days of summer, the second sub-model working days of winter and so on, the overall model is the combination of all sub-models. The concept of multiple models assumes that the data to be modelled are generated by the mixture of models and has been used successfully to minimize the effect of abrupt change of the values of the attributes of the dataset [10-12]. Clustering is combined with domain knowledge in the identification of sub-models.

The rest of the paper is organized as follows. Section 2 describes the data and methods utilized in this study. The results are presented in section 3 and are discussed in section 4. Finally, the conclusions and directions for future work are outlined in section 5.

II. Data and Methods

This section describes the dataset employed in the study, how the short-term load forecasting model was built, the method currently used by EDM to forecast electricity consumption and how to estimate DAM. The section also

978-1-4799-1276-6/13/$31.00 ©2013 IEEE

describes how the amount of electricity is converted into financial value.

A. The dataset

To build the forecast model we used information from several sources:

- Electricity consumption of Maputo – Measurements of hourly electricity consumption of the period from January 2003 to October 2012;
- Historical temperature in Maputo - The maximum and minimum daily average temperature in Maputo city for the period of January 2003 to October 2012;
- Season - Mozambique has mainly two seasons: the dry and cold season (winter) from April to September and wet and hot season (summer) from October to March;
- Type of day: Holiday, working day, Saturday or Sunday.

This information was integrated, resulting in a dataset with 29 attributes and 3592 instances. Each instance represents a single day. The attributes are date, hourly consumption (24 attributes), maximum temperature, minimum temperature, season and type of day. The dataset was split into a training set, corresponding to January 2003 to December 2011 (3287 instances) and a test set corresponding from January to October 2012 (305 instances).

To understand the data, following techniques were used:

- interaction with expert of the area of the electricity;
- Plotting (visualization) the electricity data and temperature data to determine the trends, behaviours and relation between attributes;
- Clustering the electricity data with the objective to describe them;

Both electricity and temperature data contained errors. The electricity consumption data were characterized by defective load curves due to high or low consumptions values. The defective days were replaced with the average of the past three similar days of the week, for example a defective Monday was replaced by the average of the last three Mondays. The same procedure was used for other days of the week.

The temperature data were characterized by missing values. The missing temperatures values were replaced by average temperatures of the month. For example if the minimum temperature of 15th of august was missing, the average minimum temperature of the month august was used as the temperature of 15th of august. The same procedure served to handle missing values for maximum temperatures.

The training and test set were corrected separately.

B. Short Term Load Forecasting Model

The consumption of electricity is non-linear; it changes with weather, season and type of day. This means that the forecast model designed for specific condition could not be accurate for different condition. Based on this assumption, five sub-models were designed, each one to predict electricity consumption for a specific type of the day. This approach is called multiple models [10-13]. The idea of using the multiple models is to decompose a complex system in simple sub-systems to facilitate the modelling and improve accuracy.

The five types of days are: working days of summer, working days of winter, Saturdays, Sundays and Holidays. They were identified by a combination of clustering of electricity consumption data and the knowledge of the experts of EDM. For clustering, the EM algorithm of Weka's workbench was used with the parameter number of cluster set to 3 and the electricity consumption as input data. The number of clusters was determined after several iterations.

The clustering identified three types of days: working days of summer, working days of winter and non-working days. The experts of the EDM suggested to split the cluster non-working days in three types of days: Saturdays, Sundays and Holidays. The combination of clustering and domain knowledge resulted in five types of days, early mentioned, each one corresponding to one sub-model. Other approaches for identifying the sub-models could be used, see e.g., [10-13].

Once each sub-model is composed by the same type of days, is assumed to be linear and possible to mode with linear model. The robust regression model was used to build each sub-model. Other algorithms such as artificial neural networks [14, 15] and SVMs [15] could be used to build the sub-models. The robust regression model [6] was selected due to its relative simplicity and robustness in the presence of outlier.

Equation 1 shows the demand forecasting model based on the concept of multiple models. Df is the demand to forecast, t+1 is the day to forecast, ci is the type of the day (where i is the index of the sub-model), D is the actual demand, t is the actual day, b1, b2 and b3 are the regression coefficients, T_M is the prediction of maximum temperature of the day to forecast and Tm is the prediction of minimum temperature of the day to forecast.

$$Df_{ci}^{t+1} = b_{1ci}D_{ci}^{t} + b_{2ci}T_M^{t+1} + b_{3ci}T_m^{t+1} \qquad (1)$$

Data from January 2003 to December 2011 were used to train the sub-models of equation 1 and data from January 2012 to October 2012 were used to test the models.

The accuracy of the model is evaluated by the Mean Absolute Error (MAE) given by the equation 2.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |Pk - (Dr + DAM)| \qquad (2)$$

Where N is the number of data points, Pk is the available energy, Dr is real demand and DAM is forecasted DAM. Ideally Pk must be equal to Dr+DAM.

C. Short-term load forecasting: The EDM Model

Currently EDM forecasts electricity for the next day by calculating the maximum consumption for each hour based on the previous 12 days of the week. For example, if the day to be forecasted is Monday, the 12 previous Mondays are used to find the maxim value of each hour.

D. Estimation of DAM

Equations 3 and 4, provided by EDM, describe the process of estimating the electricity to be sold in the Day-Ahead-Market (DAM). The amount of electricity to be sold to DAM should be what remain after satisfying the internal needs.

$$DAM = round(P_k - D_f) * 0.90 \qquad (3)$$

$$DAM = 0 \ \ if \ (P_k - D_f) \leq 0 \qquad (4)$$

In equation 3, Pk = 350MW is the available electricity, acquired from suppliers based on conracts. Df is the forecasted demand for the next day. round() is a function that rounds down (e.g. 3.4=3 and 2.8=2) the difference between Pk and Df . The coefficient 0.90, was defined by EDM, serves to reduce the result of the difference between Pk and Df in order to minimize the risk of committing to DAM more energy than available. Equation 4 shows that when the forecasted demand exceeds the available electricity, no energy will be sold to DAM.

E. Estimation of financial values

The financial values equivalents to the electricity available for DAM were obtained by multiplying the amount of electricity for DAM with the average market clearing price (MCP) given in USD. Because the values of MCP were obtained as monthly average, the values of DAM were aggregated per month before the multiplication. The values of MCP for the period January to October 2012 were obtained from the report of SAPP [8, 16].

The financial values equivalent to the underestimated electricity for DAM are calculated by multiplying the tariff at which the EDM would buy the electricity at that particular time of day, by the amount of electricity underestimated. The tariff varies according to season and time of day (peak hour, standard hour and off-peak hour). The season of June and August is the most expensive because has the highest demand [16]. The peak hours are the most expensive and off-peak hours are the cheapest. The tariffs used in this study were obtained from the catalogue of ESKOM for years 2012/2013 [17].

The tariff for the calculation of the financial values equivalent to the overestimation depends on the amount of electricity overestimated. If the amount overestimated is above 15% of the available electricity (Pk), the emergency tariff is used. The emergency tariff is the highest tariff and varies in accordance to the period and the time of day. If the overestimated amount is less than 15% of the available electricity, is used the same tariff that EDM would buy electricity, multiplied by two. This tariff depends also on the time of day. For this study the emergency tariff was provided by EDM.

III. Results

This section presents the results of estimating the DAM (see section D) using the two different models for demand forecasting; the model currently used by EDM (described in section C), and the new model presented in this paper (described in section B). For reference, in this paper we will call them EDM and RR (robust regression) respectively. Both models were trained using data from January 2003 to December 2011 and tested using data from January 2012 to October 2012.

Figure 1 shows the result of the automatic clustering of the electricity consumption data with weka's workbench, this result was obtained during the pre-processing with the objective to describe the electrical consumption data. The clustering of figure 1 was used by the domain expert to determine the five types of days, which were used to build sub-models of Model RR: sub-models for winter working day, summer working days, Saturdays, Sundays and Holidays.

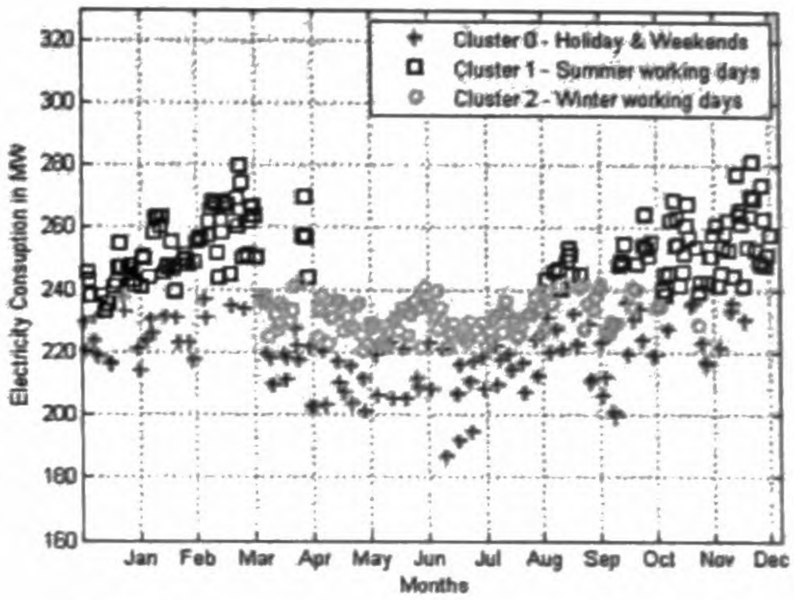

Fig.1.- Clusters of electrical consumption, showing difference between working days of winter, summer and non-working days. This result was used by domain expert to derive following five clusters: Summ working days, winter working days, Saturdays, Sundays and Holidays

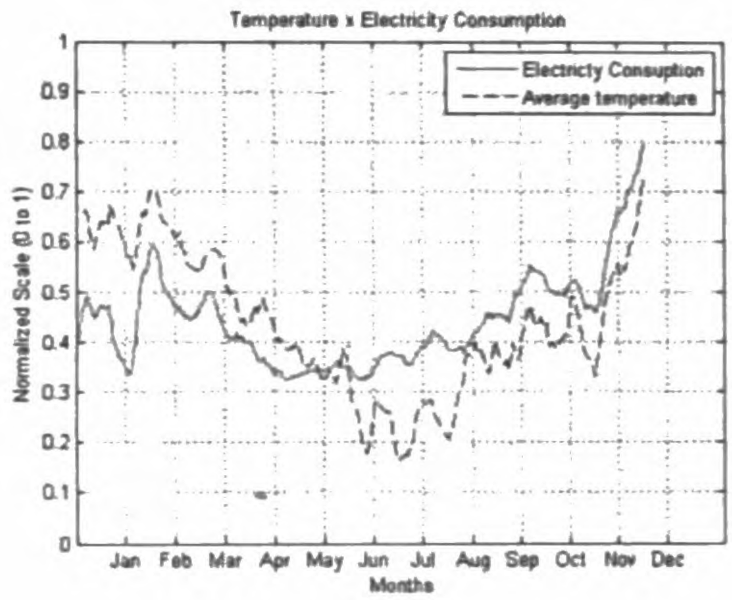

Fig.2: Comparison of the electricity consumption and temperature. There is a positive correlations between the two variables.

Figure 2 shows the comparison between the electricit consumption and the temperature during the year. It is note that the consumption of electricity increase when th temperature increases and decrease when the temperatu decrease. This correlation determined the selection of th temperature as the variable for the forecasting model.

Figure 3 presents the DAM estimated using mod EDM (DAMedm), DAM estimated using the model R (DAMrr), the real demand (Dr), the total consumption f the EDM (Dr+DAMedm), the total consumption for th model RR (Dr+DAMrr) and the available electricity (Pk The graph in Figure 3 shows only the first week of June ar was selected arbitrarily from the result of applying bo models to the test set. The stacked bar chart of figure compares the underestimated electricity by the model ED (black portion) and by the Model RR (not colored portio per month. The stacked bar chart of figure 5 compares t overestimated electricity by the model EDM (back portio and by the Model RR (not colored portion).

Table 1 presents the Mean Absolute Error (MAE) f the evaluation of Models EDM and RR respectively. Table presents the monthly revenues obtained by applying t models DAMedm and DAMrr, together with the co resulting in the overestimation by the model DAMedm a DAMrr respectively. The last column of Table 2 is t ifference between the revenues by the model DAMrr and AMedm after deducting the cost related to overestimation. or example the difference of the month of January is alculated as following:

34,371.10-247.866.84)-(2,625,799.30-127,063.00)=687,767.96.

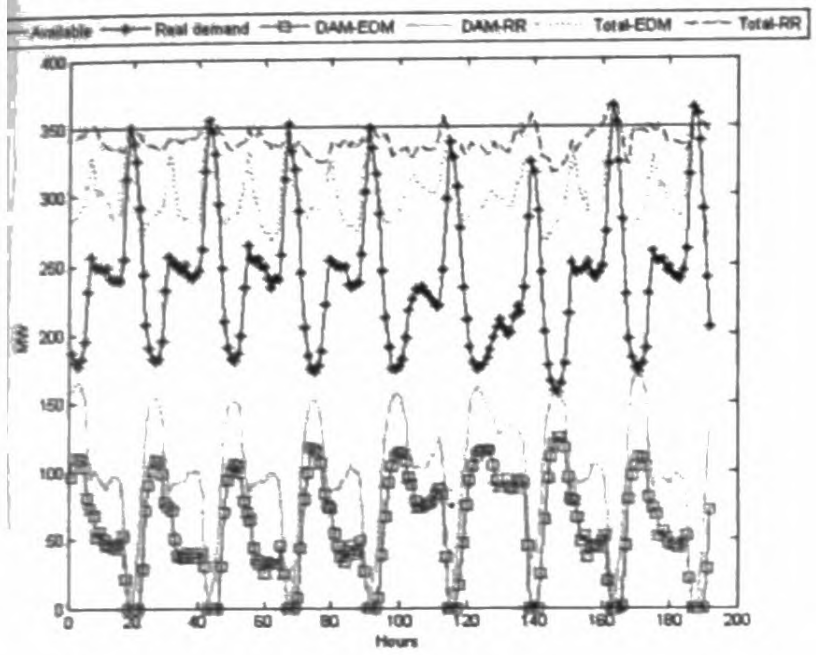

Fig.3- Comparison of the results obtained by model EDM and Model RR. The model RR loss less electricity than the model EDM.

	Model EDM	Model RR
MAE	31.20	13.02

Table 1: Mean Absolute Error (MAE) of models EDM and Model RR respectively

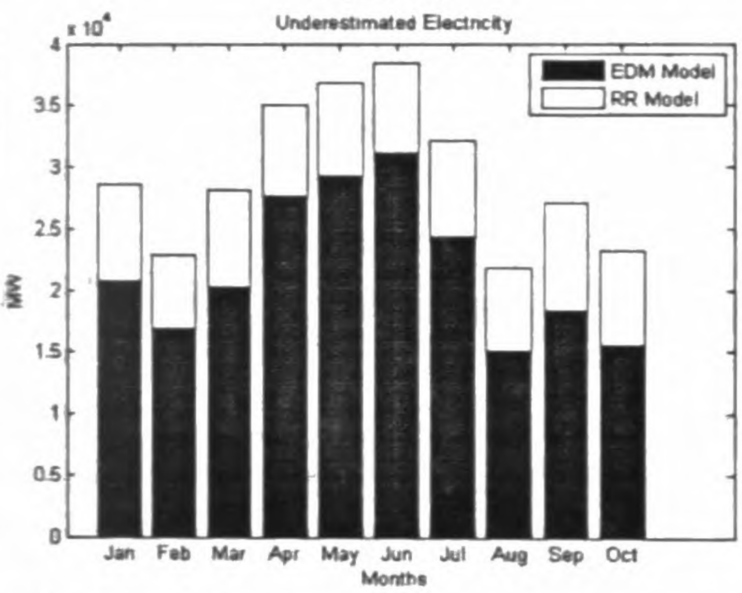

Fig. 4 - Comparison between the underestimated electricity by the odel EDM and model RR. Model RR underestimate less electricity than Model EDM

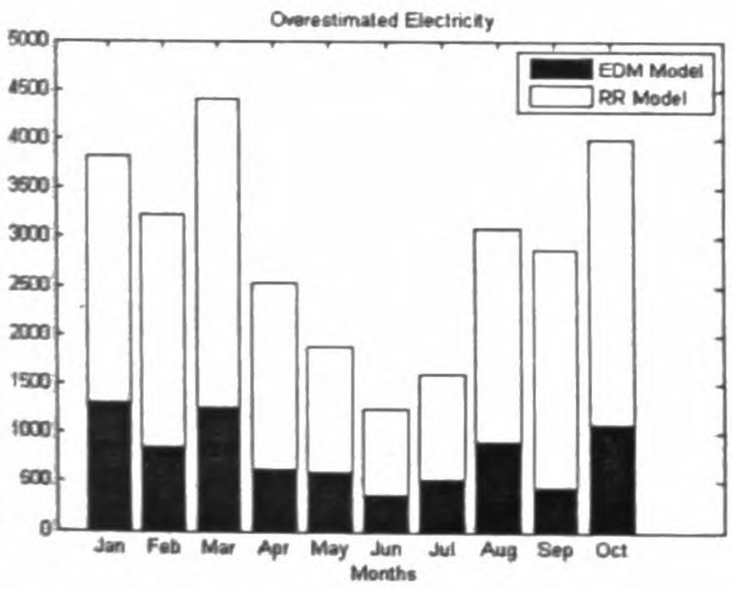

Fig. 5. Comparison between the overestimated electricity by the model EDM and Model RR. Model RR Overestimate more energy than Model EDM.

Month	Revenue DAMedm	Cost OVEedm	Revenue DAMrr	Cost OVErr	Diference
jan	2,625,799.30	127,063.00	3,434,371.10	247,866.84	687,767.96
feb	2,369,497.20	112,581.43	3,101,807.60	223,413.99	621,477.84
mar	1,970,771.30	173,993.23	2,636,892.20	342,749.96	497,364.17
apr	2,671,978.80	92,806.40	4,044,773.20	211,835.41	1,253,765.39
may	3,052,263.80	88,934.37	4,629,429.50	139,700.00	1,526,400.07
jun	3,063,529.90	178,699.01	4,711,075.00	236,965.64	1,589,278.47
jul	4,831,463.40	250,340.74	6,356,750.60	332,323.02	1,443,304.92
aug	4,683,128.40	311,088.03	5,399,073.10	492,586.49	534,446.24
sep	2,701,306.40	60,413.27	3,248,548.20	260,537.72	347,117.35
oct	1,766,400.50	129,504.08	2,093,872.30	305,015.08	151,960.80
Tot	29,736,139.00	1,525,423.55	39,656,592.80	2,792,994.15	8,652,883.20

Table 2: Financial implication of the models EDM and Model RR in USD

IV. Evaluation and Discussion

The results show that the RR model is better than the EDM model. The RR model allocates more electricity for DAM and reduces the amount of electricity lost due to underestimation. From table 2 it is observed that the application of the RR model increase the revenues in USD 8,652,883.20 in relation to the EDM model during the period January 2012 to October 2012.

In Fig. 3, it is noted that the area between the total consumption of the RR model and the available electricity is smaller than the area between the total consumption of the EDM model and the available energy. This means that the RR model is able to forecast the DAM with more accuracy than the EDM model. The curacy of the RR model is also confirmed by the results of the MAE. The MAE of RR model is less than of EDM Model as shown in table 1.

In Fig. 4, is observed that the RR model overestimates more electricity for DAM than the EDM model. This result is not desirable, because the overestimation is expensive, since the electricity for replacement is purchased at high price because is non-planned purchase. For the present case, the effect of this unwanted result is small and is compensated by the good estimation of DAM by the RR model. The results presented in table 2 show that the good estimation of DAM by RR model compensates the slight increase of overestimation of electricity.

The use of clustering to describe the data, provided insight to the domain expert, and facilitated the selection of type of days for the construction of sub-models. This is an example of how data mining can contribute to better decision making.

F. Socio-economic implications

The good accuracy obtained by the model RR contributes with socioeconomic benefits. The first benefit is for the electricity company which increases the revenue by selling more electricity and reduces the loss due to the lower underestimation. The increase of income of the company enhances its capability to invest in electrical infrastructure to serve more efficiently and effectively. The expansion of the electricity infrastructure is a key objective in Mozambique, because despite a huge potential in energy resources, only 18% of the population has access to electricity [18]. The reduction of losses can potentially influence the reduction of electricity prices for end-users, making it more affordable.

The result also contributes to the efforts to reduce adverse impact on the environment created by the use of firewood based energy and charcoal due to limited access and high cost of electricity. Cuvilas et al [19] indicate that 81 % of the energy used in Mozambique is wood fuel and there are increasing use of charcoal both in rural and in urban area which contributes to the deforestation.

One of the implications of the results is the development of price policies and awareness campaigns to promote the use of electricity more rationally and encourage more use of electricity during the time of day when it is cheaper. This type of actions can reduce the over-consumption of energy during peak hours and make more reserves for DAM.

G. Limitations

Despite the above reported advantages, the model RR has the drawback of increasing the level of overestimation. Compared to the overall contribution of the model RR, the effect of the increase of the overestimation is small; nevertheless, improvement of the model RR is desired to reduce this drawback and to increase its accuracy.

V. Conclusion and Future Work

This study presents a new short-term load forecast model for the city of Maputo, Mozambique. The model is compared with the one currently used by the EDM. The results shows that the new model (RR) is more accurate than the one used by EDM, allowing for more electricity to be allocated to DAM, which reduces the loss of electricity due to underestimation. The model RR has the drawback of slightly increasing the amount of overestimated electricity, but this small negative effect is compensated by the increases in revenue created by the model RR. The analysis of the results indicated that the increase in the amounts of electricity to sell and the savings of electricity obtained by applying the model RR can contribute to effort for widening the access to electricity and consequently reducing adverse effects on the environment created by the use of energy from charcoal and wood fuel.

The results show how data mining can benefit organizations, in this case, the clustering method helped to determine sub-models used to build a more accurate load forecasting model which in turn served to decide the amount of electricity to sell to DAM. The results may contribute to the increase in revenue and reduction of the adverse effects on the environment. Future work includes improving the model in order to further enhance its accuracy and solve the problem of increasing the over-consumed electricity. The integration of more data such as environmental, socioeconomic and regarding the operation may help to increase the accuracy of the future forecasting models.

VI. Acknowledgements

The first author acknowledges the support of the SIDA/SAREC at UEM, the EDM in special Mr. Sandro Chiang for valuable support and advices during the development of this work and the national institute of metrology for providing the temperature data.

References

[1] E. A. Feinberg and D. Genethliou, "Load Forcasting," in *Applied Mathematics for Restructured Electric Power System: Optimization, Control and Computational Inteligence*, J. H. Chow, *et al.*, Eds., ed Nwe York: Springer Science + Business media, 2005, pp. 269-285.

[2] H. Yu-Jun, et al., "Similar day selecting based neural network model and its application in short-term load forecasting," in Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on, 2005, pp. 4760-4763 Vol. 8.

[3] C. Ying, et al., "Short-Term Load Forecasting: Similar Day-Based Wavelet Neural Networks," Power Systems, IEEE Transactions on, vol. 25, pp. 322-330, 2010.

[4] M. B. Stojanovic, *et al.*, "Mid-Term Load forecasting Using recursive Time Series Prediction Strategy with Suport vector Machines," *Facta Universitatis*, vol. 23, pp. 287-298, 2010.

[5] R. Weron, "Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach," in *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, ed: John Wiley & Sons, 2006, pp. 1-23.

[6] J. Liu and J. Yu, "Short-term load forecasting system using data mining," in *Automation and Computing (ICAC), 2011 17th International Conference on*, 2011, pp. 183-188.

[7] SAPP. (2013, 02/14). About SAPP: http://www.sapp.co.zw/viewinfo.cfm?linkid=7&siteid=1.

[8] SAPP, "SAPP Day Ahead Market (DAM) Monthly Performance report: November 2012," Harare2012.

[9] H. Mori and N. Kosemura, "A data mining method for short-term load forecasting in power systems," *Electrical Engineering in Japan*, vol. 139, pp. 12-22, 2002.

[10] V. Cherkassky and Y. Ma, "Multiple model regression estimation," *Neural Networks, IEEE Transactions on*, vol. 16, pp. 785-798, 2005.

[11] G. Gregorcic and G. Lightbody, "Local Model Network Identification With Gaussian Processes," *Neural Networks, IEEE Transactions on*, vol. 18, pp. 1404-1423, 2007.

[12] K. Narendra and Z. Han, "A new approach to adaptive control using multiple models," *International Journal of Adaptive Control and Signal Processing*, vol. 26, pp. 778-799, 2012.

[13] K. S. Narendra and J. Balakrishnan, "Adaptive control using multiple models," *Automatic Control, IEEE Transactions on*, vol. 42, pp. 171-187, 1997.

[14] C. Zhang and P. Ma, "Short-Term Electricity Price Forecasting Based on PSO Algorithm and RBF Neural Network Algorithm," in *Measuring Technology and Mechatronics Automation (ICMTMA), 2010 International Conference on*, 2010, pp. 334-337.

[15] A. Selakov, *et al.*, "A comparative analysis of SVM and ANN based hybrid model for short term load forecasting," in *Transmission and Distribution Conference and Exposition (T&D), 2012 IEEE PES*, 2012, pp. 1-5.

[16] SAPP, "Anual Report 2012," SAPP Coordination Centre, Harare2012.

[17] ESKOM, "ESKOM: Tarifs & Charge Booklets 2012/2013," ESKOM, Ed., ed. Johanesburg, 2012, p. 21.

[18] G. Mahumane, *et al.*, "Energy Outlook for Mozambique 2012-2030: LEAP-Based Scenarios for Energy Demand and Power generation," presented at the Mocambique: Acumulacao e Transformacao em Contexto de crise Internacional, Maputo, 2012.

[19] C. A. Cuvilas, *et al.*, "Energy situation in Mozambique: A review," *Renewable and Sustainable Energy Reviews*, vol. 14, pp. 2139-2146, 2010.