

# Prediction of Tanzanian Energy Demand using Support Vector Machine for Regression (SVR)

Baraka Kichonge  
Nelson Mandela African  
Institution of Science  
and Technology (NM-  
AIST)  
P.O. Box 447 Arusha,  
Tanzania

Geoffrey R. John  
College of Engineering  
and Technology (CoET),  
University of Dar es  
Salaam (UDSM),  
P.O. Box 35131 Dar es  
Salaam, Tanzania

Thomas Tesha  
College of Informatics  
and Virtual Education  
(CIVE), University of  
Dodoma (UDOM),  
P.O. Box 490 Dodoma,  
Tanzania

Iddi S.N. Mkilaha  
College of Engineering  
and Technology (CoET),  
University of Dar es  
Salaam (UDSM),  
P.O. Box 35131 Dar es  
Salaam, Tanzania

## ABSTRACT

This study discusses the influences of economic, energy and environment indicators in the prediction of energy demand for Tanzania applying support vector machine for regression (SVR). Economic, energy and environment indicators were applied to formulate models based on time series data. The experimental results showed the supremacy of the polynomial-SVR kernel function and the energy indicators model in providing the transformation, which achieved more accurate prediction values. The energy indicators model had a correlation coefficient (CC) of 0.999 as equated to 0.9975 and 0.9952 with PUKF-SVR kernels for economic and environment indicators model. The energy indicators model closeness of predicted values as compared to actual values was the best as compared to economic and environment indicators models. Furthermore, root mean squared error (RMSE), mean absolute error (MAE), root relative squared error (RRSE) and relative absolute error (RAE) of energy indicators model were the lowest. Long-run sustainable development of the energy sector can be achieved with the use of SVR-algorithm as prediction tool of future energy demand.

## Keywords

Energy demand, energy demand indicators, energy prediction, support vector machine for regression

## 1. INTRODUCTION

Notwithstanding its extremely vivacious importance to all human activities and life in general, energy prediction studies using the machine learning approach in the developing countries like Tanzania has not been done deeply. In addition, the energy availability and the concern to its scarcity due to the depletion threat of fossil fuel resources, has made the analysis of energy demand to be of great interest to researchers. In fact, energy is important to all human activities and thus a socio-economic development catalytic agent for individual and nations in general. The energy analysis using various approaches for different applications has assisted individuals and countries to plan for their energy demands ahead of time. Tanzania is among the developing countries where intensive investments are taking up in all sectors of the economy. The country energy demand is expected to grow [1] as new investments floods in due to economic sectors expansions and liberalization especially in gas, minerals and agriculture. To facilitate and assist energy policy makers in decision-making, this study adopts support vector machine for regression (SVR) to analyze the influence of economic, energy and environment indicators in the prediction of energy demand of Tanzania. The choice of SVR is due to its strong computational capabilities. SVR has previously used for a number of applications such as electricity load forecasting [2,

3]; predicting crude oil price [4]; wind speed estimation [5]; classification [6, 7]; among many others. Expectations are the study results to presents an effective tool for the prediction of long-term energy demand based on time series data.

## 2. SUPPORT VECTOR MACHINE

The support vector machines (SVMs) in machine learning are supervised learning models with associated learning algorithms that analyze data and recognize patterns [8]. SVMs are applicable for classification and regression analysis. When SVMs are used for classification they involves identifying to which of a set of categories a new observation belongs, on the basis of a training set of data containing observations whose category membership is known [9]. SVM for regression applies a loss function to solve various regression problems; and it has contributed to a broad range of problems arising in various fields. It is a training algorithm for learning regression rules from data which can be used to learn linear-SVR, polynomial-SVR, RBF-SVR and PUKF-SVR [10]. PUKF-SVR has been demonstrated to work well with approximation of the linear, polynomial-SVR or RBF-SVR feature space. It has further been shown to really act like linear, polynomial-SVR or RBF-SVR [10]. The detailed theory of SVM is well given in [8, 11-14] and the theory of kernels in [15]. An overview concept of SVR and PUKF-SVR function is as presented in this paper.

### 2.1 Support Vector Machine for Regression (SVR)

Support vector regression (SVR) is an SVM version for regression [8, 16]. The scholars Schölkopf, et al. [13] and Üstün, et al. [10] approach SVR by considering a data set  $[(x_1, y_1), \dots, (x_n, y_n)]$  (d-dimensional input space) and  $y$  in  $R$  space, basically arguing that, SVR tries to find the function  $f(x)$ , which relates the measured input object (say, for this case energy indicators) to the desired output property of this object (say, predicted energy demand value in MTOE) as represented in equation 1. The variables  $W$  and  $b$  represents the slope and offset of the regression function. The solution for this regression problem is solved by minimizing equation 2.

$$f(x) = WX + b \quad (W, X \in R^d) \dots \dots \dots (1)$$

$$\frac{1}{2} \|W\|^2 + C \sum_{i=1}^n L_{\epsilon}(f(x_i), y_i) \dots \dots \dots (2)$$

Where  $C > 0$  and

$$L_{\epsilon}(f(x_i), y_i) = 0 \text{ if } |y_i - f(x_i)| \leq \epsilon \text{ and } L_{\epsilon}(f(x_i), y_i) = |y_i - f(x_i)| - \epsilon \text{ otherwise } \dots \dots \dots (3)$$

$\frac{1}{2} ||W||^2$  as given in equation 2, is the term characterizing the model complexity (flatness) whereas C is the regularization constant which determines the trade-off between the model complexity  $f(x)$  and the amount up to which deviations larger than  $\epsilon$  are tolerated [11-14]. Large values of C favor solutions with few errors and small values denote preference towards low-complexity. The reformulation of equation 2 by introduction of the slack variable  $\xi_i$  and  $\xi_i^*$  gives the primal equation 4 which refers to the formulation of the regression problem in the original data space [17].

The primal formulation of the problem is suitable in case the number of objects is (much) larger than the number of involved variables; otherwise, the so-called dual is used. The slack variables  $\xi_i$  and  $\xi_i^*$  are introduced in the situation that the target value (property of the input object) exceeds the numerical limits of the  $\epsilon$  tube. The points outside the  $\epsilon$  tube are named support vectors and in fact are the vectors supporting the actual regression model [10]. The support vectors machine contribute only to building the regression function whereas the rest of the input data in the space are not important and can be rejected after the regression model is built. This is termed as sparsely of the solution where only a few data from the input space are actually taken into account in building the regression function. Therefore we get at the formulation of the approximation function as stated in [18].

Minimize

$$\frac{1}{2} ||W||^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \text{ subject to } y_i - \langle w, x_i \rangle - b \leq \epsilon + \xi_i \langle w, x_i \rangle + b - y_i \leq \epsilon + \xi_i^* \text{ and } \xi_i, \xi_i^* \geq 0 \dots \dots (4)$$

Finally, intuitively taking into consideration of the non-linear regression by including the mapping to the feature space, equation 1 can be re-constructed into equation 5 by introducing the Lagrange multipliers.

$$f(x) - \sum_{i=1}^n (\alpha_i - \alpha_i^*) \langle \phi(x_i), \phi(x) \rangle b \dots \dots \dots (5)$$

In equation 5, the model parameters  $\alpha_i$  and  $\alpha_i^*$  represent the Lagrange multipliers satisfying the constraint  $0 < \alpha_i, \alpha_i^* < C$ . These parameters can be obtained by maximizing the dual formulation, which can be derived from equation 4.

$$\text{Maximize } -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle \phi(x_i), \phi(x_j) \rangle + \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i - \epsilon \sum_{i=1}^n (\alpha_i - \alpha_i^*) \dots \dots \dots (6)$$

$$\text{Subject to } \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \text{ and } \alpha_i - \alpha_i^* \in [0, C] \dots \dots \dots (7)$$

According to Cristianini and Shawe-Taylor [14], with the Karush–Kuhn–Tucker conditions, it is only a small number of coefficients  $\alpha_i$  and  $\alpha_i^*$  will be nonzero, and the data points associated with these parameters are referred to the support vectors of the model. The vector inner product  $\langle \phi(x_i), \phi(x) \rangle$  in equations 5 and 6 represent the mapping function from the input space to feature space. These can be replaced by the generic kernel function  $K(x_i, x)$ . The kernel function represents the underlying relationship between the input data and the output property (desired output) to be modeled. Therefore modifying equation 6 by introducing the kernel function it becomes equation 8.

$$f(x) - \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b \dots \dots \dots (8)$$

As a result, the kernel function transforms the nonlinear input space into a high dimensional feature space in which the solution of the problem can be represented as being a straight linear problem.

## 2.2 Kernel Idea

Kernel-based algorithms operation idea is to transform the data in the input space into a high dimensional Hilbert space [10, 19, 20]. That is to say, a space spanned by inner-product or distance based functions of real-valued vectors representing physical entities which is referred to as the corresponding feature space [10]. In this way, it becomes possible to solve the problem as if the feature space was linear separable. Over these two decades intensive kernels based methods for SVR have been studied, proposed and the field is now in its maturity stage [17, 21]. The linear, polynomial-SVR and RBF-SVR represented in equations 9-11 respectively, are well implemented and tested in the SVR. Furthermore, the kernel based on PUKF-SVR has been implemented and tested. A detailed explanation of the PUKF-SVR is well covered in [10] and the following section gives a brief discussion.

$$K(x_i, x_j) = (\langle x_i, x_j \rangle + 1) \dots \dots \dots (9)$$

$$K(x_i, x_j) = (\langle x_i, x_j \rangle + 1)^d, \quad d = 2 \text{ Poly\_1} \dots \dots \dots (10)$$

$$K(x_i, x_j) = e^{\left(\frac{-\|x_i - x_j\|^2}{2\delta}\right)} \dots \dots \dots (11)$$

when  $\delta = 3$  the equation is polynomial – SVR 2 (Poly\_2)

$$\delta = 8 \text{ Poly\_3}$$

$$\delta = 0.5 \text{ RBF-SVR}_1$$

$$\delta = 2 \text{ RBF-SVR}_2$$

## 2.3 Pearson VII Universal Kernel (PUKF-SVR)

PUKF-SVR was proposed by Karl Pearson in 1895 and it is a special case of Type IV (symmetrical) of the families of distribution he proposed after noting that not all distribution had distributions that resembled the normal distribution [22]. The general form of the Pearson VII function for curve fitting purposes is as given in equation 12 [10].

$$f(x) = \frac{H}{\left[1 + \left(\frac{2(x-x_0)\sqrt{2^{1/w}-1}}{\delta}\right)^2\right]^{2/w}} \dots \dots \dots (12)$$

From equation 12, H is the peak height at the centre  $x_0$  of the peak, and  $x$  represents the independent variable. The parameters  $\delta$  and  $w$  control the half-width (also known as Pearson width) and the tailing factor of the peak. The main reason to use the Pearson VII function for curve fitting is its flexibility to change, by varying the parameter  $w$ , from a Gaussian shape (when  $w$  approximates infinity) towards a Lorentzian shape ( $w$  equal to 1) as depicted in Fig. 1 [10]. The function was nominated to be used as the kernel because of its flexibility to vary between a Gaussian and a Lorentzian shape and beyond. This property makes it able to serve as a kind of universal kernel which can replace (by selecting the appropriate parameter setting) the set of commonly applied kernel functions, such as the linear, polynomial-SVR and RBF-SVR kernels. The PUKF-SVR function is tested to be a valid kernel functions because its matrices belongs to the class of the symmetric and positive semi-definite matrix, which is a requirement for any function to be a kernel. The Pearson function in equation 12 is modified to suit the kernel in equation 13.

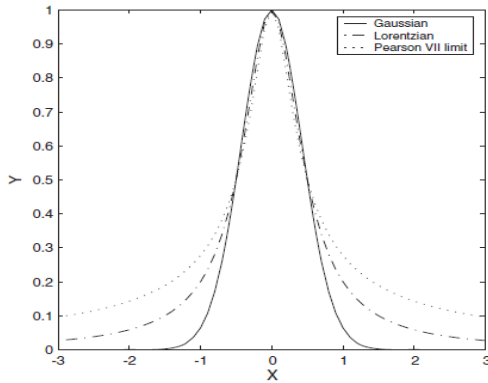


Fig 1: Pearson VII peak shapes [10]

As the Pearson width  $x=1$ , it resembles a Lorentzian and as it approaches infinity, it becomes equal to a Gaussian peak shape. Moreover, a Pearson peak with  $x=0.5$  is shown in Fig. 1. Note: the region close to zero can be imagined to compare to RBF-SVR and higher order polynomial-SVR function shapes. The region between 0.25 – 0.75 represents a linear function. Extremely, the full range between 0 – 3 becomes more or less comparable to a sigmoid function, which is widely used in neural network modeling.

$$K(x_i, x_j) = \frac{1}{\left[1 + \left(\frac{2(\|x_i - x_j\|^2)\sqrt{2^w - 1}}{\delta}\right)^2\right]^w} \dots \dots (13)$$

As can be envisaged, the single variable  $x$  in equation 12 is replaced by two vector arguments and the Euclidean distance measure between these vectors has been introduced. The peak offset term  $x_0$  is removed and the peak height  $H$  is simply replaced by 1, this without loss of generality.

### 3. METHODOLOGY

#### 3.1 Data Collection and Preprocessing

Data used were from National Bureau of Statistics (NBS), World Development Indicators, International Energy Agency (IEA), Bank of Tanzania (BoT) and Tanzania Electric Supply Company Limited (TANESCO). The dataset had the historical annual data over the period from 1990 to 2011. The dataset included population, gross domestic product (GDP), per capita energy use, total primary energy supply, gross national income per capita, electricity generation and CO<sub>2</sub> emissions. The pre-processing of the data to fit in the models was done. The three models based on the indicators of study were economic, energy and environment. The models were developed with the objectives of determining the influence of indicators in the prediction of energy demand.

#### 3.2 Experimental Setup

In the experiment, SVR was used for the study. The training to build the regression model used for evaluation involved the polynomial-SVR, normalized polynomial-SVR, RBF-SVR and the PUKF-SVR kernels. Data for all the experiments were cross-validated using  $k$ -folds cross-validation (CV). The idea was to split the data into  $k$  disjoint and equally sized subsets. The validation was done on a single subset and training was done using the union of the remaining  $k-1$  subsets. This procedure was repeated  $k$  times, each time with a different subset for validation. The intention was to allow for the large data in the dataset to be used for training and all cases appear for the validation cases (testing). For this case, the true errors were estimated as the average error rate.

### 3.3 Performance Evaluation

The models' performances in both approaches were compared and evaluated using an appropriate choice of the following statistical parameters: correlation coefficient (CC) [23], root mean squared error (RMSE), mean absolute error (MAE), root relative squared error (RRSE) and relative absolute error (RAE). The values of statistical indices were derived from statistical calculation of observation in the models output predictions and are given in Armstrong and Collopy [24] and Chattefuee and Hadi [25]. Selection of the appropriate kernel and the accurate model for prediction of energy demand was done by considering the combination of higher CC and the lowest RRSE, RMSE; MAE and RAE values.

## 4. RESULT AND DISCUSSION

To demonstrate the SVR capability on energy prediction, three experiments were conducted using the cross-validation with 10 folds for the training data. The first experiment involved the economic, the second energy and the last one environmental indicator. The value for  $k$  was experimentally chosen to be 10 folds; and thus the union of 9 folds were used for the training and the remaining subset for validation set (testing) in each cycle of one experiment.

### 4.1 Analysis of the Kernels Performance

The results of the kernels performance analysis regarding the economic indicators model as shown in Table 1 and Fig. 2 suggests PUKF-SVR kernel performed excellently in comparison to its counterparts. It had the highest CC value of 0.9975 while the RBF-SVR kernel had the lowest CC value in that case. The PUKF-SVR kernel had the lowest MAE and RMSE values of 0.1934 and 0.2589 respectively. Furthermore, the lowest RAE and RRSE characterize PUKF-SVR kernel in relation to the other kernels. The error value findings as depicted in Fig. 2 provide the comparison of errors for the various kernels involved. The two algorithms maps achieved by the Polynomial-SVR and PUKF-SVR appeared to be slightly close in most of the years with the PUKF-SVR attaining the lower value in most cases.

Table 1: Kernels statistical performance comparison-economic indicators model

	Normalized Polynomial SVR	Polynomial SVR	RBF-SVR	PUKF-SVR
CC	0.9904	0.9912	0.4383	0.9975
MAE	0.461	0.406	2.9243	0.1934
RMSE	0.5989	0.4941	3.3271	0.2589
RAE	13.59%	11.97%	86.20%	5.70%
RRSE	15.79%	13.03%	87.73%	6.83%

The results of the kernels performance analysis on the energy indicators model for the prediction of energy demand using the normalized polynomial-SVR, polynomial-SVR, RBF-SVR and the PUKF-SVR are depicted in Table 2. The polynomial-SVR kernel had the greatest predictive ability with the correlation coefficient of 0.999. The RBF-SVR and the normalized polynomial-SVR kernels had the least CC value with the RBF-SVR having the smallest CC value of 0.4961. The MAE and RMSE values of polynomial-SVR are shown to be 0.1448 and 0.1629 respectively outperforming the other kernels. The polynomial-SVR further exhibits the lowest RAE and root relative square error vindicating it to be the better estimating or predictor of energy of energy demand under energy indicators model. These can as well be spotted in Fig.

3. Even though PUKF-SVR and Polynomial-SVR appears to have similar values over the considerable range, the predictive capability went down beyond the year 2010 making the polynomial-SVR a better approach for the prediction of energy demand for this case.

**Table 2: Kernels statistical performance comparison- energy indicators model**

	Normalized Polynomial SVR	Polynomial SVR	RBF-SVR	PUKF-SVR
CC	0.6411	0.999	0.4961	0.9977
MAE	2.544	0.1448	2.7927	0.1465
RMSE	2.8269	0.1629	3.1987	0.2552
RAE	74.99%	4.27%	82.32%	4.32%
RRSE	74.54%	4.30%	84.34%	6.73%

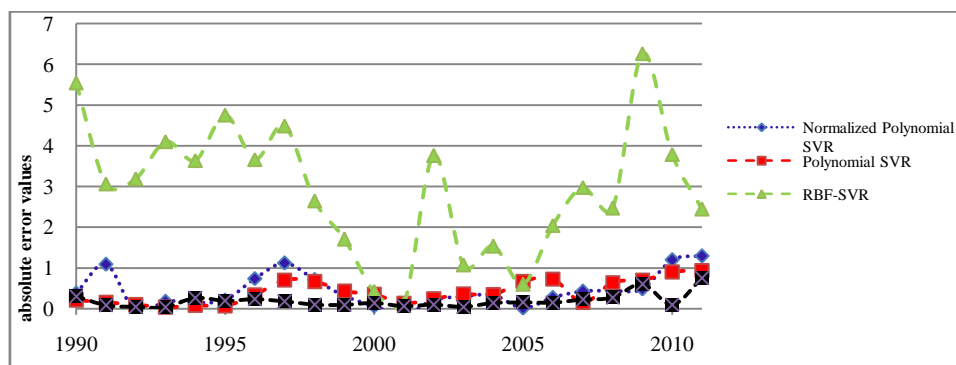
The last experiment was evaluating the effect of the kernels in the use of environment indicators model for the energy demand prediction. Table 3 shows that the greatest predictive validity algorithm was PUKF-SVR, which had the CC value of 0.9952. It is as well noted to have the lowest values for MAE and RMSE of 0.2331 and 0.3686 respectively. The RAE was 6.872% and the RRSE is 9.72. The RBF-SVR kernel had again the least CC value. The absolute errors comparison between predicted and actual values for both algorithms is illustrated in Fig. 4. The PUKF-SVR and the polynomial-SVR had slightly closer results although in most cases again PUKF-SVR values were the lowest. This puts the PUKF-SVR to be a better kernel for energy prediction using the environment indicators model.

**Table 3: Kernels statistical performance comparison- environment indicators model**

	Normalized Polynomial SVR	Polynomial SVR	RBF-SVR	PUKF-SVR
CC	0.8120	0.9934	0.4375	0.9952
MAE	1.5012	0.3323	2.8991	0.2331
RMSE	2.1296	0.4282	3.3048	0.3686
RAE	44.25%	9.79%	85.46%	6.87%
RRSE	56.15%	11.29%	87.14%	9.72%

## 4.2 Models Performance Comparison

Two visible plausible conclusions can be drawn here. The first one involves the best performing indicators model on energy demand prediction based on time series data and the second the overall better performing kernel regardless of the models.



**Fig. 2: Absolute errors comparison between kernels – Economic indicators model**

This section begins with the best performing indicators model for energy prediction. Although it is noted, the PUKF-SVR kernel had a better performance over its counterparts in both the economic and environment indicators models, thorough analysis in the energy indicators model results shows the polynomial-SVR kernel had the greatest performance over the PUKF-SVR kernel. Comparison of kernels in Table 4 shows that the polynomial-SVR has the highest correlation coefficient of 0.999 with the energy indicators model while in the economic and environment indicators models, the correlation coefficients are 0.9975 and 0.9952 respectively. Polynomial-SVR kernel for energy indicators model achieved the least values in terms of MAE as compared to the PUKF-SVR kernel in the economic and environment indicators model. The PUKF-SVR kernel had MAE values of 0.1934 and 0.2331 respectively for economic and environment indicators models. These statistical values are greater in comparison to the MAE values of 0.1448 for the energy indicators model making it the best.

**Table 4: Statistical values performance comparison**

	Economic Indicators model	Energy indicators model	Environment indicators model
SVR Kernel	PUKF-SVR	Polynomial SVR	PUKF-SVR
CC	0.9975	0.999	0.9952
MAE	0.1934	0.1448	0.2331
RMSE	0.2589	0.1629	0.3686
RAE	5.7%	4.27%	6.87%
RRSE	0.07%	4.30%	0.09%

Similarly, in terms of the RMSE values, the polynomial-SVR kernel in the energy indicators model had a lower value of 0.1629 while the PUKF-SVR kernel for both economic and environment indicators model had a higher values of 0.2589 and 0.3686 respectively. Not only these, but also RAE value and RRSE values for economic and environment indicators models are similarly higher valued as compared to energy indicators model. Furthermore, the absolute errors deviations values between actual and predicted energy demand is relatively very small for the polynomial-SVR kernel as illustrated in Fig. 3. It is further as suggested earlier that the polynomial-SVR kernels works well with the energy indicators model than is the PUKF-SVR kernel although it had shown better results with the economic and environment indicators model. These comparisons concludes that the energy indicators model were more accurate for the prediction of energy demand with the use of polynomial-SVR kernel in comparison to the economic and environment indicators models using PUKF-SVR kernel.

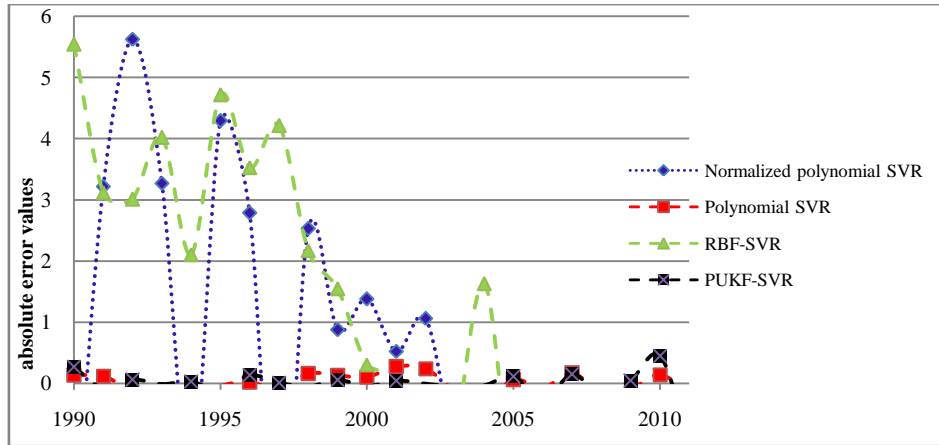


Fig. 3: Absolute errors comparison between kernels – Energy indicators model

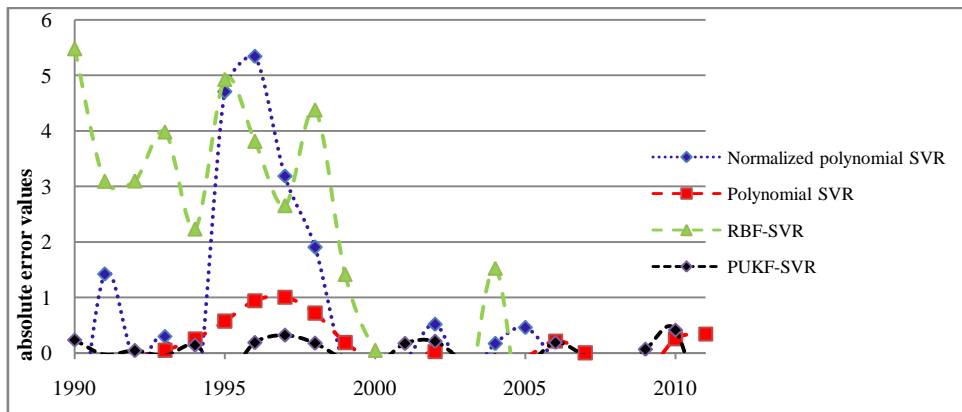


Fig. 4: Absolute errors comparison between kernels – Environment indicators model

### 4.3. Energy Prediction Curve

Figure 5 depicts the prediction of the energy demand for the period between 1990 and 2011 using the energy indicators model which emerged as a better predictor with polynomial-SVR kernel. It can be noted that the curve approximates well

the energy demand over the period of interest. This curve demonstrate the practicability of the support vector machine for regression (SVR) in the real time energy demand prediction for both short and long term.

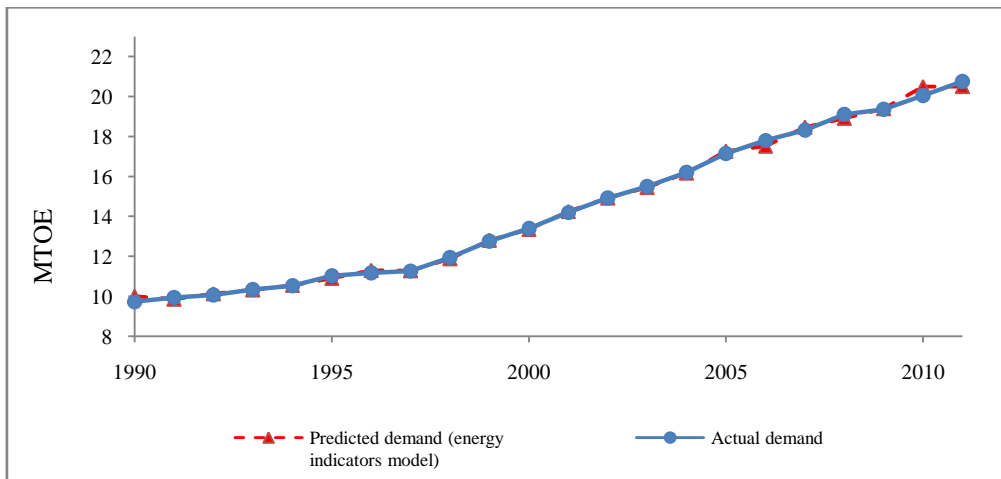


Fig. 5: Energy demand prediction curve comparison using polynomial-SVR kernel

## 5. CONCLUSION

The application of the support vector machine for regression (SVR) with normalized polynomial-SVR, polynomial-SVR, RBF-SVR and PUKF-SVR kernels functions in the analysis of energy demand was discussed in this paper. The economic, energy and environment indicators derived from time series

data were used to build the energy models. The statistical performance indices applied to evaluate the estimating ability of these techniques within SVR were correlation coefficient (CC), root mean squared error (RMSE), mean absolute error (MAE), root relative squared error (RRSE) and relative absolute error (RAE). The comparison of the experimental results to the kernel functions reveals the possibility of the use

of the SVR for the analysis and prediction of the energy demand in Tanzania. The analysis of the kernels used show that the polynomial-SVR kernel function with the energy indicators model provided the transformation, which achieved more accurate prediction values with the SVR. The use of SVR algorithm in estimating future energy demand will endeavor government in decision making on expected energy demand for the long-run sustainable development of the country. Although SVR has shown good results in the prediction of energy demand, intensive study of its comparison with other learning algorithm is of future interest. The idea is to unveil the best possible algorithm that can be implemented for the analysis and prediction of energy demand with the consideration of accuracy.

## 6. REFERENCES

- [1] Kichonge, B., John, G. R., Mkilaha, I. S. and Sameer, H. (2014) Modelling of Future Energy Demand for Tanzania, *Journal of Energy Technologies and Policy*, Vol 4, Issue No 7.
- [2] Pai, P. and Hong, W. (2005) Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms, *Electric Power Systems Research*, 74, 417-425.
- [3] Pai, P. and Hong, W. (2005) Support vector machines with simulated annealing algorithms in electricity load forecasting, *Energy Conversion and Management*, 46, 2669-2688.
- [4] Xie, W., Yu, L., Xu, S. and Wang, S. (2006) A new method for crude oil price forecasting based on support vector machines *Computational Science-ICCS 2006*, pp. 444-451 (Springer).
- [5] Mohandes, M., Halawani, T., Rehman, S. and Hussain, A. A. (2004) Support vector machines for wind speed prediction, *Renewable Energy*, 29, 939-947.
- [6] Maji, S., Berg, A. C. and Malik, J. (2008) Classification using intersection kernel support vector machines is efficient, Paper presented at the Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on.
- [7] Xue, C., Li, F., He, T. et al. (2005) Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine, *BMC bioinformatics*, 6, 310.
- [8] Olson, D. L. and Delen, D. (2008) *Advanced Data Mining Techniques* (Springer-Verlag Berlin Heidelberg).
- [9] Goel, A. (2009) Application of SVMs algorithms for prediction of evaporation in reservoirs, Paper presented at the World Environmental and Water Resources Congress.
- [10] Üstün, B., Melssen, W. J. and Buydens, L. M. (2006) Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel, *Chemometrics and Intelligent Laboratory Systems*, 81, 29-40.
- [11] Burges, C. J. (1998) A tutorial on support vector machines for pattern recognition, *Data mining and knowledge discovery*, 2, 121-167.
- [12] Smola, A. J. and Schölkopf, B. (2004) A tutorial on support vector regression, *Statistics and computing*, 14, 199-222.
- [13] Schölkopf, B., Smola, A. and Müller, K.-R. (1998) Nonlinear component analysis as a kernel eigenvalue problem, *Neural computation*, 10, 1299-1319.
- [14] Cristianini, N. and Shawe-Taylor, J. (2000) *An introduction to support vector machines and other kernel-based learning methods* (Cambridge university press).
- [15] Bishop, C. M. (2006) *Pattern recognition and machine learning* (springer New York).
- [16] Drucker, H., Burges, C. J., Kaufman, L., Smola, A. and Vapnik, V. (1997) Support vector regression machines, *Advances in neural information processing systems*, 9, 155-161.
- [17] Cortes, C. and Vapnik, V. (1995) Support-vector networks, *Machine learning*, 20, 273-297.
- [18] Vapnik, V. (2000) *The nature of statistical learning theory* (springer).
- [19] Muller, K., Mika, S., Ratsch, G., Tsuda, K. and Scholkopf, B. (2001) An introduction to kernel-based learning algorithms, *Neural Networks, IEEE Transactions on*, 12, 181-201.
- [20] Schölkopf, B. and Smola, A. J. (2002) *Learning with kernels: support vector machines, regularization, optimization, and beyond* (MIT press).
- [21] Vapnik, V. N. and Vapnik, V. (1998) *Statistical learning theory* (Wiley New York).
- [22] Lahcene, B. (2013) On Pearson families of distributions and its applications, *African Journal of Mathematics and Computer Science Research*, 6, 108-117.
- [23] Lee, R. J. and Nicewander, W. A. (1988) Thirteen ways to look at the correlation coefficient, *The American Statistician*, 42, 59-66.
- [24] Armstrong, J. S. and Collopy, F. (1992) Error measures for generalizing about forecasting methods: Empirical comparisons, *International journal of forecasting*, 8, 69-80.
- [25] Chattefuee, S. and Hadi, A. S. (2006) *Regression Analysis by Example* (New Jersey, A John Wiley & Sons, Inc., Publication ).