An Approach to Multivariable Regression Analysis of a Rooftop PV System in Households' Sector in Tirana. A Novel Approach using Energy Modeling Tool

RAIMONDA DERVISHI[1], ERJOLA CENAJ[1], LORENC MALKA[2,*]
[1]Department of Mathematics, Mathematical and Physical Engineering Faculty,
Polytechnic University of Tirana,
Tirana,
ALBANIA

[2]Department of Energy, Faculty of Mechanical Engineering,
Polytechnic University of Tirana,
Tirana,
ALBANIA

*Corresponding Author

Abstract: - Solar power prediction plays an essential role in functioning, mapping, and obtaining energy and climate goals in 2030 and beyond and contributing to real-time balancing of the power system. On the other side, electricity consumption is influenced by Heating Degree Days (HDD), Cooling Degree Days (CDD), average monthly temperature, energy management, human behaviour, architecture, orientation, and many other factors. The simulations are performed based on three consecutive energy consumption data for a typical dwelling in Tirana city. Once the base case scenario is designed, the model is validated based on the monthly electricity bills. In the base case scenario are included some energy efficiency measures (EEM) afterward the optimization of the supply side using PV and Solar Water Heating is part of our simulation, too. The focus of our work is to design a correlation between the relationship consumption and generation of electricity given as a dependent variable (Y), which is regressed with weather parameters given as independent variables (Xi) under Durbin Watson statistics. The output of the study can help designers to compile a reliable power system, better utilization of energy resources, and forecasting accuracy analysis from both sides of the energy system (demand and supply side). The tested household and EEM applied in the proposed scenario may lead to an electricity reduction level of 8655 kWh per year and 76.6 % of solar fraction is used to meet the hot water demand.

Key-Words: - Energy forecasting, Multi-Variable Regression Analyses, Photovoltaic, Durbin-Watson statistic, Sustainability, and RETScreen energy modeling tool.

Received: April 11, 2024. Revised: September 3, 2024. Accepted: October 3, 2024. Published: November 5, 2024.

1 Introduction

To domesticate clean energy solutions and cope with growing energy demand and shifting existing energy systems to 100% renewable energy systems is still very critical issue for many countries worldwide, [1]. Different approaches and energy modeling tools that incorporate advanced statistical techniques are more than a condition toward flexible energy systems, and time-stratified analysis of electricity consumption as given in the case of Turkey using a regression and neural network approach, [2]. In this regard regression analysis as an exhaustive technique in statistics that supports linking independent variables influencing a matter of economic such as price fluctuations can occasionally turn deadly as a result of the domino effect they generate [3], technical or environmental and food, wheat production issues using a quantile regression analysis as stated in the study [4] is used in many studies in the field of renewable power systems. An econometric analysis of the impact and elasticity of human resources outflow and remittance with economic growth in Ukraine proves that remittance has a direct relationship with economic activity rate, a cyclical and multiplicative relationship with the inflationary process, and the last is that an indirect effect on capital using Dynamic Regression Models, [5]. Further on modeling and energy optimization of a thermal power plant using a multi-layer perception regression method enables to estimate the relationship between input-independent and output-dependent variables, forecasts the energetic data, and highlights the benefits of the Bayesian Regularization method in the energy sector, [6]. The operation enables to fit on problems related to statisticians, economics, and engineering experts to focus on the elements that matter the most, which should be ignored, and which of them influence reciprocally when a distribution and an alignment between parameters is established. A Real-Time Price (RTP) based power scheduling scheme can be implemented effectively in Smart Grid to match supply and demand. A Smart Grid has a two-way digital communication system, and it encourages customers to actively participate in different types of Demand Response programs and can be valuable to forecasting electricity prices using the Seasonal ARIMA model and Implementing Real-Time Pricing (RTP) based tariffs in a Smart Grid as given in the study of [7]. Independent variables play an important role in various regression analyses especially if they have been provided from direct onsite measurement, statistics, and other official sources. The more accurate independent variables the more precise and better understanding of dependent variables (DR) would be carried out for long-term predictions, [8]. More specifically, in energy and climate issues considering the supply side which can be shifted and optimized in terms of renewable energy sources (RES) precise prediction using statistical analyses is necessary. Such analyses are in the crave of sizing and to concisely estimate the future potential of RES power plants that are required to meet demand. In terms of prediction, electricity suppliers are concerned with various time fractions such as hourly, seasonal, and yearly production to smooth the consequences arising from the contract with the government or other end users in Albania, [9]. In recent years, the implementation of different forecasting methods applied in the energy sector, especially for PV forecasting has become a sprightly and even more interesting research field, as project revenue is strongly affected by weather data variations in time such as solar irradiance, temperature, dust, wind speed and humidity, [10], [11]. Therefore, the possibility to predict the PV power generation and other forms of RES energy systems (based on daily, monthly, and hourly up to 24 h or even more to every 15 min resolution) can be very important for well-regulated small-scale grid-connected PV systems. Applications for residential usage, are fragile as such power systems should oversee working in a well-connected power network. The forecasting methods, in the energy sector and climate-related issues, especially those related to intermittent energy sources such as photovoltaic, wind, tidal, etc. can be classified into two groups as physical models or statistical models, [12]. In the case of using the physical modeling approach, the generation in time from a tested PV is designed and modeled by inputting weather variables, mainly radiation and temperature known as numerical weather prediction (NWP) models. Instead, the statistical forecasting approach is performed when past measured time data series are available, denominated as time series graphs, [13]. The statistical models are uncomplicated and manageable compared to physical models owing to less input data and lower computation intentions to predict generation is required. Many researchers have composed the value of time series models including Auto-Regressive (AR), integrated (I), and Moving Average (MA), which can be combined to bring forth well-known ARMA and ARIMA models, [14]. These models are the most unpretentious ones performed in the field of times series forecasting as such models use historical data points of the dependent variables, acquainted as endogenous forecasting. Exogenous forecasting employs a historical dataset of independent predictors diametrically different as in the case of using dependent variables In general, nowadays mathematics methods that enable to shape in time solar radiation as an independent variable affecting the power generation from PV are nonidentical, and may include linear regression models [15], Auto-Regressive (AR) model [16], Moving Average (MA) models, Artificial Neural Network (ANN) models, [17]. Being recognizant by the fact that besides technical advancements and technological improvements photovoltaic power generation is strongly affected by a mix of sundry weather parameters. In this case to include more than one predictor in the PV modeling process, a multivariable regression model can be used to bring to fruition the most accurate and representative prediction model. In conclusion, our study is focused on the integration of small rooftop PV plants in the household sector part of the Tirana region. The integration of onsite RES with help end-users to pay reduced energy bills and will play a critical role in reducing emissions, bringing homes towards nearly zero energy building (nZEB) in 2030 and positive energy building (PEB) in 2050. Besides the security of supply, such systems would help in job creation and redesigning urban infrastructure to unlock the potential for carbon emissions reduction, [18].

2 Multivariable Regression Analysis towards Smart Grid Systems

Photovoltaic power generation for hours or days ahead can contribute to efficient and economic use of energy resources and can allow planners to adapt better the amount of energy in a future growing demand. Furthermore, the desire for the development of electrical smart grids drives the need to have a day ahead production information to channel the energy flows, optimizing benefits by applying energy storage systems such as reducing congestion problems, arbitrage, smoothing PV generation, peak shaving and deferral in distribution and transmission power lines [19], storing and delivering it with lowest impact into the grid.

3 Simple Linear Regression

The simple linear regression model is the most elementary regression model in statistics with only one predictor X and can be written as given in Eq.1, [20].

$$Y_i = b + aX_i + \varepsilon_i \; where \; i = 1, \dots \dots \dots, n, \quad (1)$$

where Y_i, X_i reflect the values of the response and predictor variables in the test, respectively; a is called the intercept, and b is the slope of the line, while the last term ε_i belongs to error from $N(0, \sigma_\varepsilon^2)$. Using some method like l_1 regularization [2], the ordinary least squares method, which relies on minimizing the sum of square of errors $\sum \varepsilon_i^2$. For the simple linear regression model the ordinary least squares estimations of a and b can be carried out by using expression in Eq. 2:

$$\hat{b} = \hat{Y} - \hat{a}\hat{X_\imath} \quad (2)$$

Or in the extended form b value can be carried out by using expression in Eq.3:

$$\hat{a} = \frac{\sum_i (X_i - \hat{X}) \cdot (Y_i - \hat{Y})}{\sum_i (X_i - \hat{X})^2} \quad (3)$$

The goodness R^2 of fit is defined from Eq.4:

$$R^2 = \frac{\sum_i (\hat{Y}_i - \hat{Y})^2}{\sum_i (Y_i - \hat{Y})^2} \quad (4)$$

3.1 Multiple Linear Regression Model

In our case study, the power generation from PV plant in time and weather variables were used as an input to frame a multiple regression model with interaction effects, which occurs when the effect of the predictors varies. Such interaction is gifted as a product of the diverse predictors (independent weather variables) representing an extension of the general regression model as given in Eq.5.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots . \beta_n x_n + \varepsilon \quad (5)$$

β_0 represents the intercept and each x belongs to various independent weather variables (predictors), while the other β parameters belong to the slope coefficient of the variable and are defined in the designed model with the focus to minimize the error ε. As a result, the interaction effects of the multiple regression model can be written (Eq.6).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \cdots . \beta_n x_n + \beta_1 x_m x_n + \varepsilon \quad (6)$$

$x_1 x_2$ represents the interaction between two variables.

3.2 Matrix Form of Multiple Regression

The matrix form of multiple regression model consists of a column of ones and p column vectors of the observations on the independent variables and can be written as in Eq. 7.

$$M = \begin{pmatrix} 1 X_{11} \; X_{12} \cdots \; X_{1p} \\ 1 X_{21} \; X_{22} \; \dots X_{2p} \\ \vdots \; \vdots \quad \vdots \; \vdots \; \vdots \\ 1 X_{n1} X_{n2} \; 1 \; X_{np} \end{pmatrix} \quad (7)$$

The vector of parameters to be estimated using formula in Eq.8

$$\beta \in R^{p+1} \quad (8)$$

on the other hand, β can be carried out using Eq. 9.

$$\beta = (\beta_0, \beta_1 \dots \dots \dots \dots . . \beta_p)^T \quad (9)$$

for ε values taken in Eq.10

$$\varepsilon \in R^n \quad (10)$$

$$\varepsilon = (\varepsilon_1, \varepsilon_2 \dots \dots \dots \dots . . \varepsilon_n)^T \quad (11)$$

The vector β belongs to a vector of unknown constants and is matter of estimation from the data by $\hat{\beta}$. The normal form can be written as following:

$$X^T X\hat{\beta} = X^T Y \tag{12}$$

If $X^T X$ has an inverse, then the unique solution of normal equations given by Eq. 13:

$$\hat{\beta} = (X^T X)^{-1} \cdot X^T Y \tag{13}$$

The vector $\hat{Y}$ of estimated means of the dependent variable Y for the values of the independent variables $X_1, X_2 \dots\dots\dots\dots.. X_n$ in the dataset is computed as given in Eq. 14:

$$\hat{Y} = X\hat{\beta} \tag{14}$$

However, to express $\hat{Y}$ as a linear function of Y expression in Eq.15 can be used.

$$\hat{Y} = \lfloor X(X^T X)^{-1} \cdot X^T \rfloor Y \tag{15}$$

4 The Energy Modeling Tool

4.1 RETScreen Expert

Almost always the system's complexity, available energy data, demographic issues, economic and policy factors, the way of accessing and real-time technical support, all-in-one concept execution, and the country context are the main factors influencing the right energy modeling tool, [1]. Driven by the above-mentioned factors, the RETScreen energy model, a clean energy management software platform that enables low-carbon planning, implementation, monitoring, and reporting reliably [20] as the modeling tool is chosen. The supply side (PV generation), and demand side (tested household energy consumption) in the light of different energy efficiency measures, locations, and technical systems including the optimization of the supply side can be carried out. Various climate data are provided by the RETScreen Climate Database (CanmetENERGY), which has integrable Energy Resource Maps in a high-resolution way, [21]. The model flowchart from base case scenario up to proposed scenario and regression analyses is given in Figure 1.

The multiple combinations regression analysis fits in creating multiple regression analysis for multiple combinations of various predictors in one sickle grabbing. By alleviating a bond between considerable factors of influence and a dependent variable such as the demand side and supply side of energy (consumption/production) with a single mathematical equation and evaluating the most influencing drivers, [20]. The model is very highly sweeping and can be used besides all for educational and teaching purposes [1], such as the University of Toronto, Polytechnic University of Tirana, etc.

Fig. 1: Energy model flowchart supported by RETScreen Expert: Modified after [1], [20], [21]

4.1.1 Application of Multivariable Regression Model with Multiple Combinations of Factors in Household Sector

After all, the selection of independent variables (e.g. factors of influence), in a regression that predicts the consumption side or supply side known as the dependent variable, sags on technical or other characteristics (age, size, geometry, construction, and location) that impact the overall energy system from both sides demand side and supply side, [1]. In our case study climate parameters are the main and crucial factors of influence in the demand side section. The consumption of residential buildings is splited to meet HVAC and DHW and many other services and equipment (lighting, refrigerators, cooking, cleaning, etc.). Consumption is impacted highly by weather-dependent variables including daylight, humidity, HDD, CDD, temperature, etc. In the study demographic characteristics, population age, and economic aspects are not considered [22], as different categories of home users may require higher energy demands and higher comfort levels due to income level. Climate could also be responsible for short-term fluctuations in energy consumption, as milder than usual weather could decrease annual energy demand, while severe winter or hot summer seasons could cause consumption peaks, as our national annual energy balance is characterized, [23]. Hence, in principle, better monitoring of energy use in dwellings can be achieved by correcting the demand side to counterbalance the effects of weather regressed with HDD/CDD or temperature, while photovoltaic generation can be regressed with horizontal solar radiation, tilted solar radiation, and average temperature. This way of portraying our energy system enables the creation of a multivariable regression, which in the end can be merged [20] including the beginning date and duration of the baseline scenario. First, the baseline or reference period is constructed using three annual consecutive energy consumption from the electricity bills and power production of the tested PV plant. The last step is the choice of the regression model that can be employed. There are two options of choosing daily or duration, [20]. In our analyses, a "Daily" method is used as it is given in Table 1. The daily regression method facilitates the creation of a daily average for each dependent variable, and the coefficient calculated by the model, while "Duration" uses the time intervals that experts evaluate of importance, and no averaging is used, [20]. In Table 1 the independent variables that impact consumption and production for the tested dwelling located in Tirana city are given.

Table 1. Representation of the dependent function (Y) and independent variables used to regress the demand and supply side of the tested household

Dependent variable	*Electricity Consumption (kWh)*	*Y*	*Variable Order*
Independent variable	HDD (°C-d)	X_1	1
Independent variable	Average Monthly Temperature (°C)	X_2	2
Independent variable	CDD (°C-d)	X_3	3
Method			Daily
Weighted			Yes
Dependent variable	Electricity generation (kWh/month)	Y	Variable Order
Independent variable	Daily solar radiation – horizontal ($kWh/m^2/d$)	X_1	1
Independent variable	Daily solar radiation – tilted ($kWh/m^2/d$)	X_2	2
Independent variable	Average temperature (°C)	X_3	3
Method			Daily
Weighted			Yes

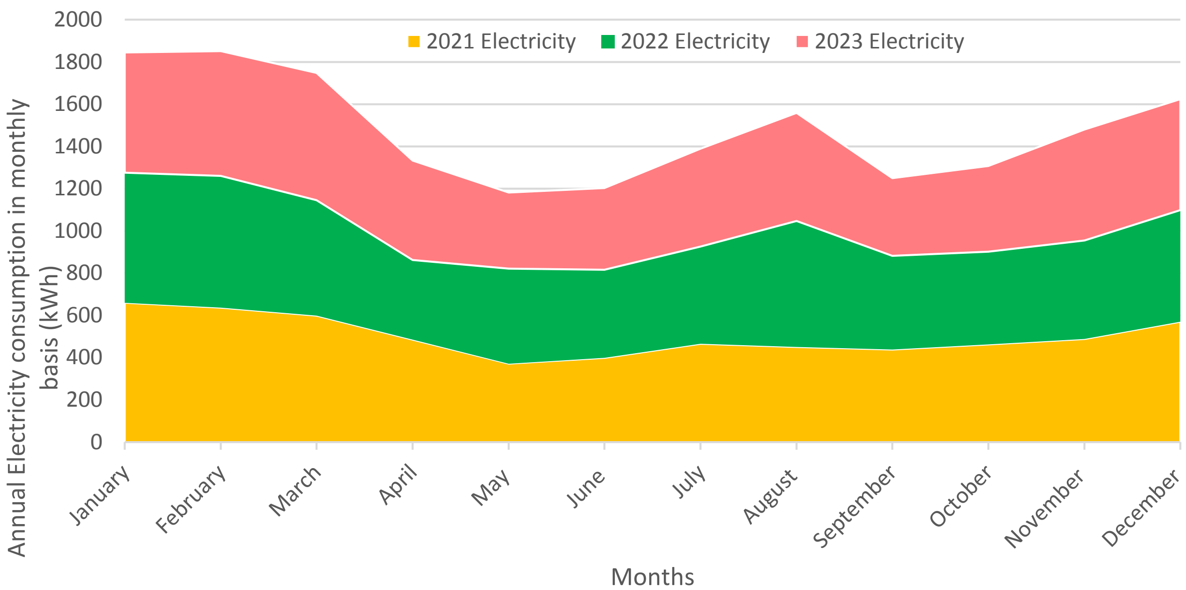

Fig. 2: Historical electricity consumption balance (2021 – 2023) of the tested household (kWh)

The energy consumption referring to the last three years is provided from the owner's official electricity bills. In Figure 2 the electricity bills are given monthly basis for the following years 2021, 2022, and 2023. The average electricity consumption for the chosen period falls between 490-500 kWh per month. The peak demand for electricity is observed in the winter months, January, February, November, and December.

As can be seen from Figure 2 electricity demand is not equally distributed on a monthly and yearly basis, too. Based on these data and forecasted weather data taken one can find of important to develop a correlation between energy consumption as a dependent variable and weather parameters such as HDD, CDD, and monthly average temperature known as independent parameters. Durbin Watson's statistic of the chosen variables to detect the presence of autocorrelation at lag 1 in the residuals from our regression analysis is used. The Durbin-Watson equation can be given from the expression in Eq. 16 and can be served as one way to test to determine whether autocorrelation is present, [20].

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T}(e_t)^2} \tag{16}$$

The results and interpretation of the regression will also change if other predictors are added into the chosen control volume. The method, in cases when predictors are missing, may affect the regression, in that case, the idea to include as many predictors as possible does not work due to six reasons explained in the study, [1].

5 Simulation and Results

In this study, the regression of both sides of our energy system is performed. The regression uses three following electricity data consumption (Figure 2) defined as the demand side used as primary values to be regressed as a function of HDD as given in Figure 3. The total number of observations is 36, each month represents an observation value that should be regressed.

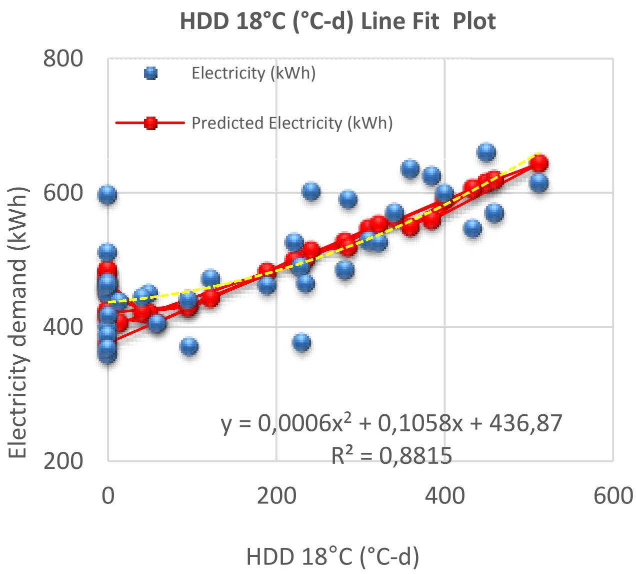

Fig. 3: Correlation between electricity demand and Heating Degree Days (HDD) (°C-d) at 18ºC

Figure 3 is given the prediction of electricity demand (kWh) as a function of HDD. The simulation result it is evidenced a good correlation between electricity demand for space heating and HDD leading to an R^2 value of 0.8815. The mathematical relation has polynomial form and is represented by the Eq. 17:

$$y = 0.0006x^2 + 01058x + 436.87 \quad (17)$$

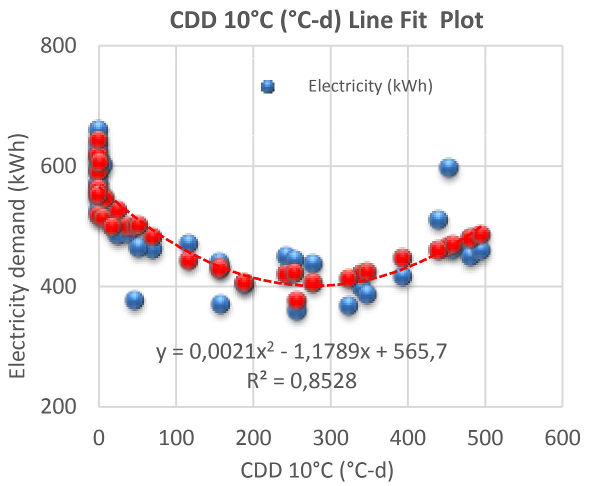

Fig. 4: Correlation between electricity demand and Cooling Degree Days (CDD) at 10ºC (°C-d)

In the graph in Figure 4 the prediction of electricity demand (kWh) as a function of CDD for the tested household is given. In this case, R^2 is lower than regressed with HDD (Figure 3) and results 0.8528. The expression that correlates electricity demand as a function of CDD is represented in Eq. 18:

$$y = 0.0021x^2 - 1.1789x + 565.7 \quad (18)$$

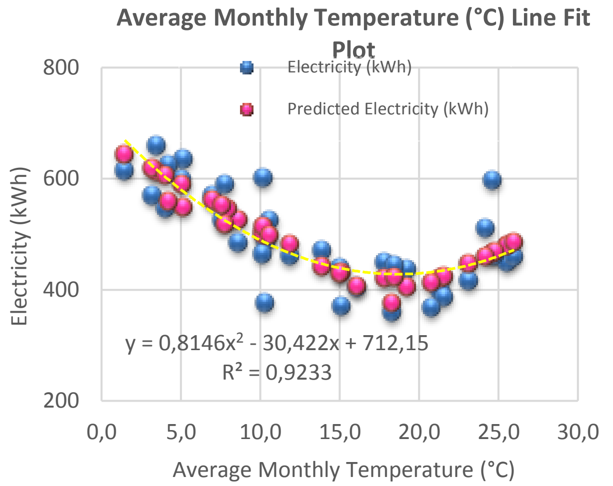

Fig. 5: Correlation between electricity demand and Average Monthly Temperature (°C)

In the graph in Figure 5 the prediction of electricity demand (kWh) as a function of Average Monthly Temperature (°C) is plotted. The simulations show that R^2 is higher than values carried out in the graphs in Figure 3 and Figure 4. In this case, a goodness fit value R^2 of 0.9233 is carried out. The mathematical relation of electricity consumption as a function of Average Monthly Temperature as ploted in Figure 5 has polynomial form and can be represented from Eq. 19:

$$y = 0.8146x^2 - 30.422x + 712.15 \quad (19)$$

Based on the above information we can make a forecast of electricity consumption based on the above numbers. The prediction is executed between two confidence bounds, lower and upper as plotted in the graph in Figure 6.

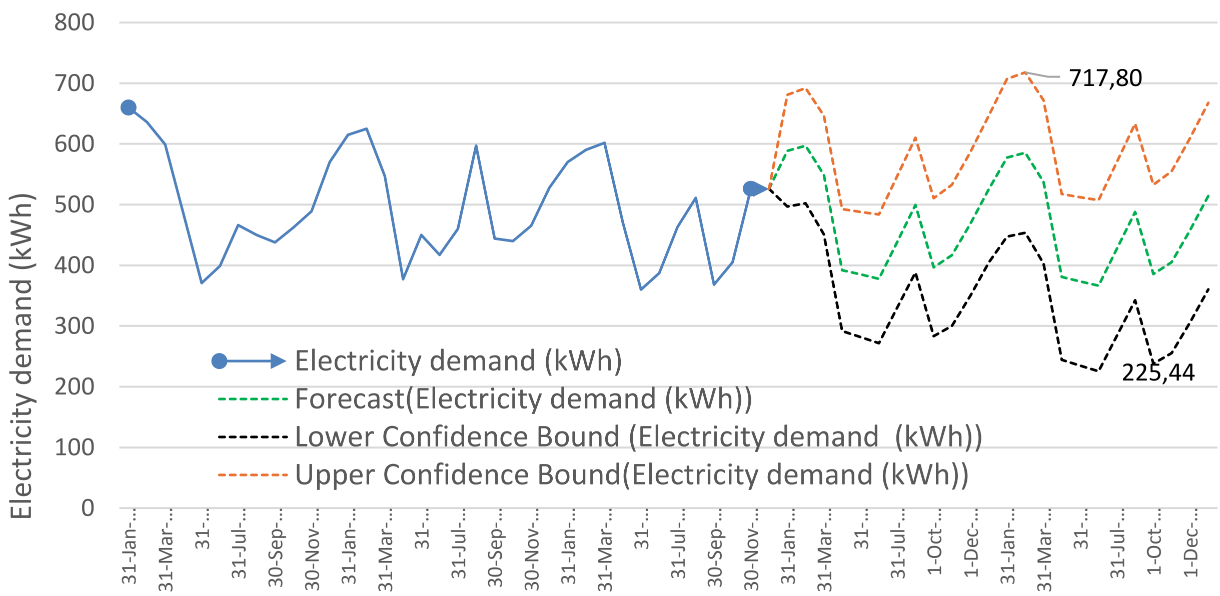

Fig. 6: Electricity consumption prediction for the tested household

To forecast the electricity consumption, there are used monthly electricity bills for three consecutive years in the tested and validated household too. While predictors (independent variables) are provided from the RETScreen weather data base. As can be seen from the graph in Figure 6 the consumption reaches a minimal and maximal value of 225.4 kWh in the lower confidence bound and 717.8 kWh in the upper confidence bound, respectively.

In the graph in Figure 7 normal probability plot of electricity demand (kWh) is given.

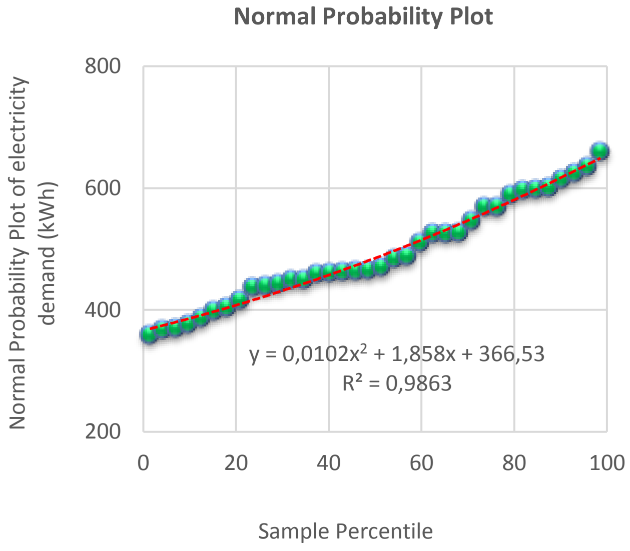

Fig. 7: Probability dispersion (%) of electricity demand (kWh) for the tested household

The normal probability distribution of electricity consumption can help decision makers to rapidly assess its' variability in the sample percentile as it is given in the graph in Figure 7.

In the graph, in Figure 8 the predicted electricity generation from a tested rooftop photovoltaic system with an installed capacity of 5.5 kW as a function of Daily solar radiation - tilted (kWh/m^2/d) is given.

As it is shown in the graph in Figure 8 the correlation between electricity generation and daily solar radiation - tilted (kWh/m^2/d) is very strong as R^2 results 0.9553. The predictor and dependent variable are linearly correlated as given in Eq.20:

$$y = 142.26x + 19.555 \tag{20}$$

In the graph given in Figure 9, the predicted electricity generation from a tested rooftop PV plant with an installed capacity of 5.5 kW as a function of average monthly temperature (ºC) is given.

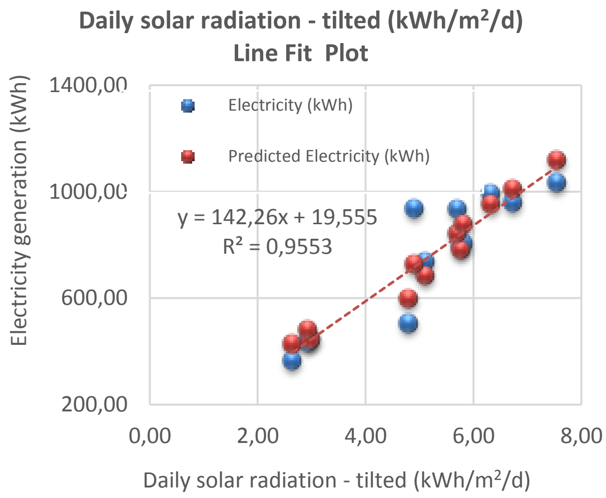

Fig. 8: Electricity generation from tested PV with a capacity of 5.5 kW as a function of Daily solar radiation - tilted (kWh/m^2/d)

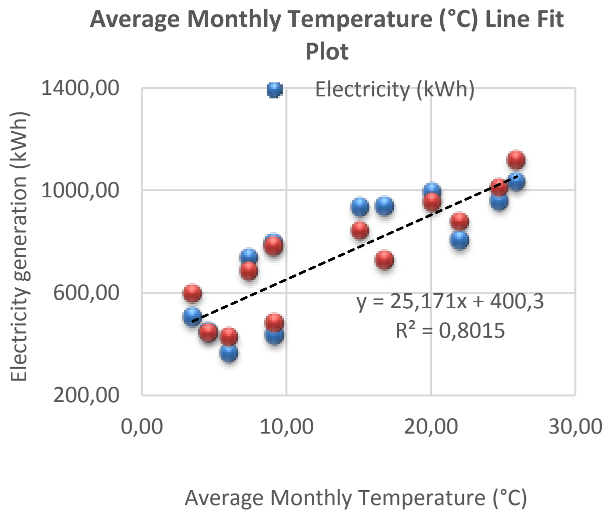

Fig. 9: Electricity generation from tested PV with a capacity of 5.5 kW as a function of Average Monthly Temperature (°C)

As shown in the graph in Figure 9 the correlation between electricity generation and Average Monthly Temperature (°C). is lower if the predictor would be Daily solar radiation - tilted (kWh/m^2/d) as given in Figure 8. In this case, R^2 results 0.8015. The predictor and dependent variable are linearly correlated as given in Eq.21:

$$y = 25.171x + 400.3 \tag{21}$$

The correlation between predicted electricity production and as a function of horizontal daily radiation (kWh/m^2/d) is given and represented by the graph in Figure 10. In this graph the correlation between predicted electricity production and as a function of horizontal daily radiation (kWh/m^2/d) is given.

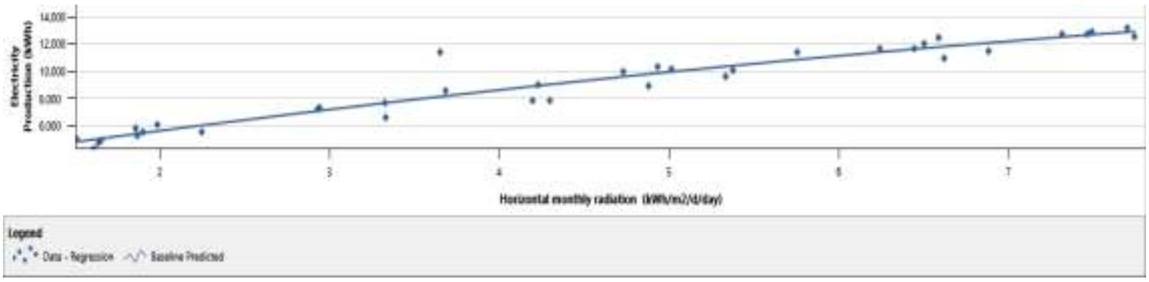

Fig. 10: Predicted electricity generation (kWh) from tested rooftop PV system with an installed capacity of 5.5 kWp as a function of horizontal daily radiation (kWh/m^2/d)

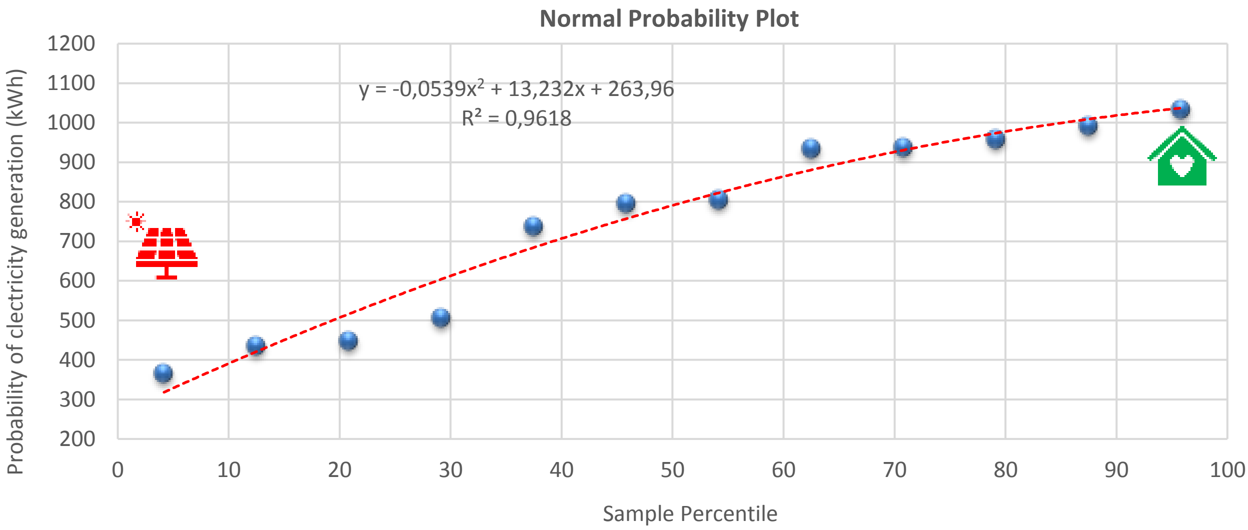

Fig. 11: Probability dispersion (%) of electricity generation (kWh) from tested rooftop PV system with an installed capacity of 5.5 kWp

The regression leads to a polynomial form of DR and independent variable as given in Eq. 22. This equation correlates horizontal monthly radiation and electricity generation (DR) for the tested the rooftop PV plant with an installed capacity of 5.5 kW.

$$y = -61.54x^2 + 1870.39x + 2115.45 \qquad (22)$$

The correlation is almost strong as the R^2 value is approximately 0.930. The above expression given in Eq.22 can be used by other experts in the field of energy planning to forecast future electricity generation from on-site PV.

The graph in Figure 11 provides a distribution of the possible values for the energy production for the tested rooftop PV plant with a capacity of 5.5 kWp. The value of each point represents the frequency (%) of values that fall in the range defined by the proposed system. As it can be seen from graph 11 only 40 % of the cases will produce less than 600 kWh per month and the rest fall between 700 and 1100 kWh per month. Analyzing the simulation result represented in the graph in Figure 11 a good balance between predicted energy demand and supply side is placed. From the simulation, an R^2 value of 0.9618 results, and the linkage between DR and independent variable is of polynomial form as given in the expression in Eq.23:

$$y = -0.0539x^2 + 13.232x + 263.96 \qquad (23)$$

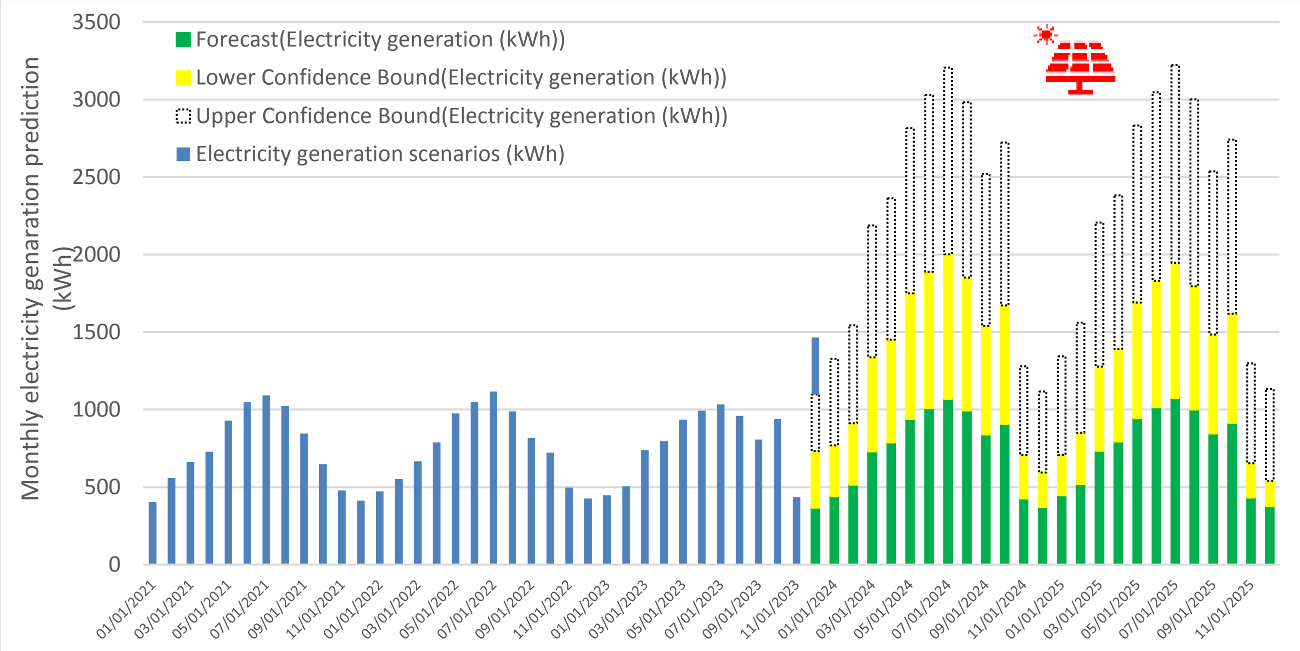

Fig. 12: Expected electricity generation trend from tested rooftop PV system with an installed capacity of 5.5 kWp

To forecast the electricity generation, there are used three consecutive electricity generations of a tested rooftop PV plant with a total installed capacity of 5.5 kWp provided by the study, [1]. In the graph in Figure 12 for the chosen PV technology and capacity a forecasting methodology using RETScreen Expert model using upper and lower confidence bound for electricity generation is carried out. The simulation of the proposed PV plant will guarantee stable electricity generation (green region graph).

6 Model Validation: Tested Household

The private apartment chosen for this case study is situated in Tirana city and is part of mid-rise category, with a living area of 115 m^2 and serves to accommodate 4 people. The energy consumption carried from the model is validated based on monthly electricity bills, identified as a base case scenario. After implementing EEM such as exterior wall insulation, double glass windows, efficiency management and replacement of existing lights to more efficient lights and performing optimization of the supply side that includes technical systems such as 5.5 kWp PV plant installed on the roof PV and solar heating water panel (SHW) with a gross area of (3.73 - 4.1) m^2, the proposed scenario is designed using RETScreen modeling tool. The proposed scenario is designed based on recommended values for heat transfer coefficients provided from [1], [24], [25]. Exterior walls of the building envelope should have a heat transfer coefficient equal to or less than 0.38 W/m^{2}K, while windows should have a heat transfer coefficient not more than 2.2 W/m^{2}K (if new construction 2.0 W/m^{2}K), [24]. Existing water heating is provided from an electric boiler (2.0 kW) with a total volume of 80 Liter. Daily water consumption is calculated from the model based on a selected occupancy rate of 100% at a desired water temperature of 60°C.

The total energy demand in the base case results in 23,132 kWh per year. The distribution of energy consumption per each end-use for the tested household is given in the graph in Figure 13.

After implementing the energy efficiency measures, the total energy consumption is reduced in the base case resulting in 14,436 kWh per year as given in the graph in Figure 14.

The distribution of energy consumption per each end-use is given in the graph in Figure 14. As can be seen from differences between graphs 13 and 14 the energy consumption is reduced by 37.6%.

In the graph in Figure 15 simulation results for the case of improved household (after EEM implementation) including optimization of the supply side for the tested household are depicted.

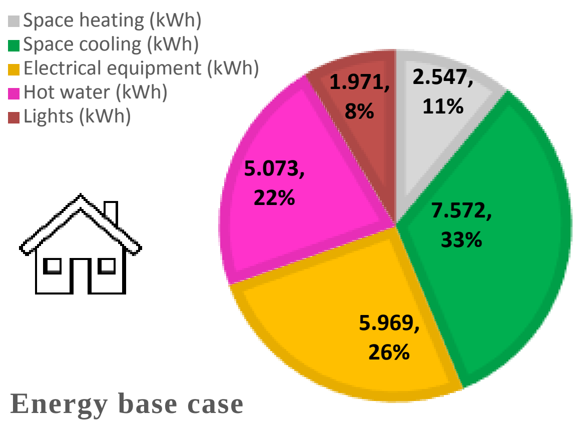

Fig. 13: Basecase scenario energy distribution by end use for the tested household

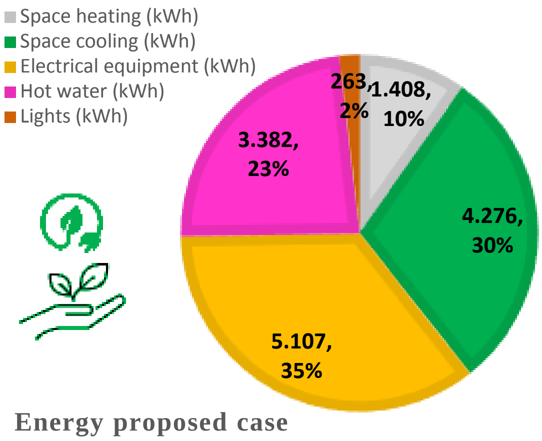

Fig. 14: Proposed case energy distribution by end use for the tested household

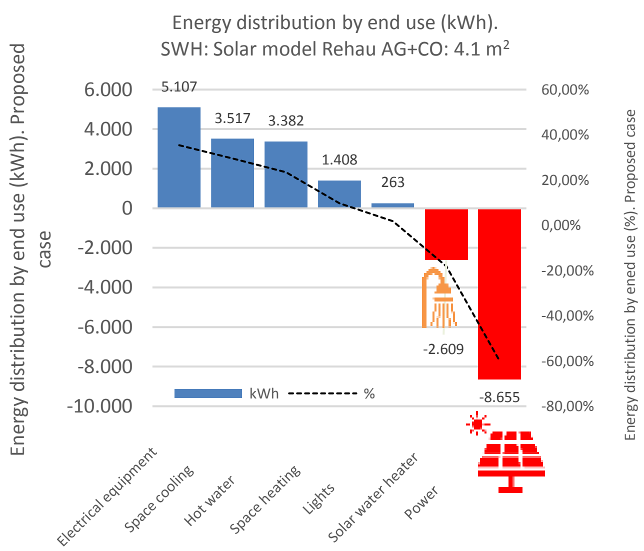

Fig. 15: Scenario Energy distribution by end use (kWh). SWH: Solar model Rehau AG+CO: 4.1 m^2

Besides energy reduction from the application of the above-mentioned EEM, when a 5.5 kWp rooftop PV power plant for power generation and a solar water heating plant (SWH) to meet hot water demand is installed it will reduce further electricity consumption by 11,264 kWh in annual basis. Hence, the proposed energy systems will provide a lot of energy to the system and if better managed it can be a very good approximation to convert the tested household to PEB as given in the study of [1] and can deliver the excess during peak off-peak hours to the national distribution line or it can be stored, using batteries, EV or it can be used to run Heat Pumps or produce H_2

7 Multivariable Regression Analyses Results and Discussion

One way to test to determine whether autocorrelation is present in a time-series regression analysis [20] is by using the Durbin-Watson test for the chosen parameters. The simulation results on the demand side (electricity consumption) as a dependent function (Y) are executed as a function of HDD, Average Monthly Temperature) and CDD as given in Table 2. As can be seen from the Table 2 the number of observations used is 36 (36 months of electricity bills). The coefficient of multiple determination (R^2) results in 0.608, while the coefficient of multiple determination – Adjusted (Ra^2) results in 0.571. F-test (p-value) for the multiple regression analyses results 1.19E-06. The Durbin -Watson statistic results in 1.3. This number is important if the results are more than 1 and less than 2 [20]. As a result, good modeling means that the proposed model can be applied for extended scientific work and long-term energy scenarios.

In that case, the correlation is not strong enough as the coefficient of multiple determination (R^2) is approximately 0.608 if the combination among independent variables is fixed as follows: X1=HDD, X2=Average Monthly Temperature, and X3=CDD. Coefficient values of a, b, c, and d are given at the end of the simulation results (Table 2) In this case the mathematical expression given in Eq.24.

$$Y = 0.035449 \cdot X1 + 19.7538 \cdot X2 - 0.00572 \cdot X3 + 1.925 \quad (24)$$

The same philosophy of multiple regression on the supply side that includes a rooftop PV plant for electricity generation as a dependent function (Y) regressed as a function of three independent weather values:

Table 2. Multivariable regression results for the case of electricity consumption using the Durbin-Watson statistic test

Variables					
Y	Electricity consumption (kWh)				
X1	HDD				
X2	Average Monthly Temperature				
X3	CDD				
Method	Daily				
Weighted	Yes				
Regression results					
Number of observations:	36				
Number of iterations:	11				
The sum of residuals:	0.876425997				
Average residual:	0.024345167				
Residual sum of squares - Absolute:	122.3769247				
Residual sum of squares - Relative:	119.2796493				
Standard error of the estimate:	1.930670619				
Coefficient of multiple determination(R^2):	0.608005179				
Coefficient of multiple determination – Adjusted (Ra^2):	0.571255664				
Root-mean-square error (RMSE):	1.95557636				
Coefficient of variation of the RMSE:	0.120510784				
F-test (p-value):	1.19E-06				
Net determination bias error (NDBE):	0.0015				
Durbin-Watson statistic:	1.4				
Coefficient results					
Equation Y=a·X1+b·X2+c·X3+d	Standard error		t-ratio	p-value	User-defined
a	0.035441774	0.00101	35.03	0	0.0354
b	19.7538	0.52840	37.38	0	19.753
c	-0.00572	0.0010	-5.364	6.89E-06	-0.0057
d	1.925	0.2741	7.0262	5.78E-08	1.9259

X1=Daily solar radiation tilted, X2=Horizontal monthly radiation and X3=Average Monthly Temperature as depicted in Table 3.

As can be seen from the simulation result in Table 3.

The number of observations for the tested PV plant is 36 The coefficient of multiple determination (R^2) results in 0.951, while the coefficient of multiple determination – Adjusted (Ra^2) results in 0.947. F-test (p-value) for the multiple regression analyses results 1.19E-06. The Durbin - Watson statistic results approximately 1.902. which is more than 1 and less than 2. As a result, a good modeling output, which means that the proposed model can be applied for extended scientific work and long-term energy scenarios.

From the simulation, it is observed that a good correlation between Y and X1, X2 and X3 is evidenced and given by the mathematical expression in Eq. 25 The coefficient of multiple determination (R^2) results 0.951 and adjusted (Ra^2) of 0.947. Coefficient values of a, b, c, and d are given at the end of the simulation results (Table 3)

$$Y = 803.85 \cdot X1 + 399.9401 \cdot X2 + 94.619 \cdot \mathrm{X3} + 1834.180 \tag{25}$$

Equation 25 can be used to approximate the electricity generation in each system of independent variables (X1, X2, and X3). In the graph, in Figure 16 combined multivariable regression for actual electricity generation versus predicted electricity generation of the proposed PV plant with a total installed capacity of 5.5 kW is depicted. As clearly shown from the graph in Figure 16 a good fit between actual electricity generation and predicted values is observed, in the case of a multivariable regression.

The regression analyses clearly show that predicted baseline electricity generation is strongly correlated with the actual electricity values as their respective lines fitted well in time. Electricity generation from tested PV plants is highly influenced by X1=Daily solar radiation tilted, X2=Horizontal monthly radiation, and X3=Average Monthly Temperature.

Table 3. Multivariable regression results for the case of electricity generation using Durbin - Watson statistic

Variables					
Y	Electricity Generation (kWh)				
X1	Daily solar radiation - tilted				
X2	Horizontal monthly radiation				
X3	Average Monthly Temperature				
Method	Daily				
Weighted	Yes				
Regression results					
Number of observations:	36.000				
Number of iterations:	12.000				
The sum of residuals:	178.980				
Average residual:	4.972				
Residual sum of squares - Absolute:	13914332.921				
Residual sum of squares - Relative:	13771108.977				
Standard error of the estimate:	656.009				
Coefficient of multiple determination(R^2)	0.951				
Coefficient of multiple determination – Adjusted (Ra^2)	0.947				
Root-mean-square error (RMSE)	659.411				
Coefficient of variation of the RMSE:	0.073				
F-test (p-value)	0.000				
Net determination bias error (NDBE)	0.001				
Durbin-Watson statistic	1.902				
Coefficient results					
Name	Value	Standard error	t-ratio	p-value	User-defined
a	803.8547792	410.6688	1.957428	0.0590	803.8547
b	399.9404001	330.8295	1.208902	0.2355	399.9404
c	94.61995227	23.90696	3.957841	0.0003	94.61995
d	1834.180139	746.6704	2.456479	0.0196	1834.180
Equation:	a·X1+b·X2+c·X3+d				

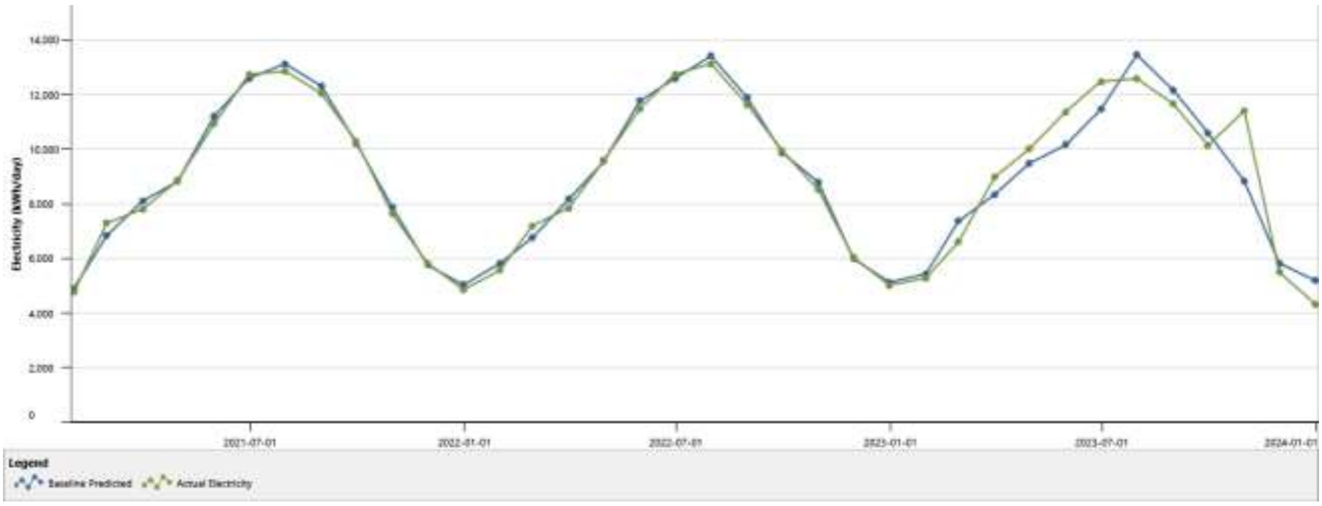

Fig. 16: Combined multi-variable regression for actual electricity generation versus predicted electricity generation from the tested rooftop PV plant with a capacity of 5.5 kWp

The coefficient of multiple determination (R^2) results 0.951, while the coefficient of multiple determination – Adjusted (Ra^2) results 0.947. The model can be extended easily on a national level as it reduces energy consumption by 8655 kWh per year and 76.6 % of the solar fraction is used to meet the water heating demand and represent a cheap option of integrating onsite RES in the residential sector in Albania.

8 Conclusion

According to the simulation executed in the RETScreen energy modeling tool is evidenced a good regression between a dependent variable and independent variables. To verify the goodness of the correlation Durbin-Watson test is used. This statistical test is used to detect the presence of autocorrelation (a relationship between energy consumption and generation) in the residuals [20] (prediction errors) from a regression analysis of the dependent function concerning independent variables indexed as X1, X2, and X3. It is observed that electricity consumption is influenced by Heating Degree Days, average temperature, and Cooling Degree Days. The coefficient of multiple determination (R^2) results in 0.608, while the coefficient of multiple determination – Adjusted (Ra^2) results in 0.571. On the supply side, electricity generation is highly influenced by X1=Daily solar radiation tilted, X2=Horizontal monthly radiation, and X3=Average Monthly Temperature. The coefficient of multiple determination (R^2) results in 0.951, while the coefficient of multiple determination – Adjusted (Ra^2) results in 0.947. In both cases, demand and supply side it is observed that Durbin – Watson statistic values are higher than 1 and less than 2. The model can be extended and may be easily applicable to analyze the effects of EEM on a national level. The tested household and EEM applied in the proposed scenario may lead to a reduction level of 8655 kWh per year and 76.6 % of solar fraction is used to meet the hot water demand. The proposed mathematical model can apply to regions with similar weather conditions, especially in the Mediterranean region as stated in [1].

9 Future Work

In the future, more variables of influence such as economic income level and other social issues that impact energy consumption within residential sectors will be included. The way toward zero emission building (ZEB) is the focus of our future research employing advanced statistical methods.

References:

[1] Malka, L., Dervishi, I., Owolabi, A. B., Abdullahi, Y., Suh, D., Aliaj, B., Zeqo, M., & Dervishi, R. (2024). Deep retrofitting toward net zero energy building in the Mediterranean region: A case study of Albania. *Heliyon*, 10(15), e35174. https://doi.org/10.1016/j.heliyon.2024.e35174

[2] Yiğit, S., Turgay, S., Cebeci, Ç., & Kara, E. S. (2024). Time-Stratified Analysis of Electricity Consumption: A Regression and Neural Network Approach in the Context of Turkey. *WSEAS Transactions on Power Systems*, 19, 96–104. https://doi.org/10.37394/232016.2024.19.12

[3] Konomi, I., & Zani, B. (2023). Forecasting Inflation using the ARIMA Approach (Case of Albania). *WSEAS Transactions on Business and Economics*, Vol. 20, pp. 1252–1259. https://doi.org/10.37394/23207.2023.20.111

[4] Hussain, B., Ali, U., Shaheen, S., & Almas, L. K. (2022). Climate Volatility, Wheat Productivity and Food Security: A Quantile Regression Analysis. *WSEAS Transactions on Environment and Development*, Vol. 19, pp.1196–1209. https://doi.org/10.37394/232015.2023.19.109

[5] Mulska, O., Vasyltsiv, T., Voznyak, H., Mitsenko, N., Kaplenko, H., & Ivaniuk, U. (2024). International Remittances, Human Resources Outflow & Economic Growth: Dynamic Regression Models. *Financial Engineering*, Vol. 2, pp.23–33. https://doi.org/10.37394/232032.2024.2.3

[6] Xezonakis, V., & Ntantis, E. L. (2023). Modeling and Energy Optimization of a Thermal Power Plant Using a Multi-Layer Perception Regression Method. *WSEAS Transactions on Systems and Control*, Vol. 18, pp.243–254. https://doi.org/10.37394/23203.2023.18.24

[7] Joshi, H., Pandya, V., Bhavsar, C., Shah, M. (2016). "Forecasting Electricity Price Using Seasonal Arima Model and Implementing RTP Based Tariff in Smart Grid," *WSEAS Transactions on Power Systems*, Vol. 11, pp. 43-51.

[8] Erten, M. Y., & Aydilek, H. (2022). Solar

Power Prediction using Regression Models. *International Journal of Engineering Research and Development*, 14(3), pp.333-342. https://doi.org/10.29137/umagd.1100957

[9] Malka, L., Kuriqi, A., & Haxhimusa, A. (2022). Optimum insulation thickness design of exterior walls and overhauling cost to enhance the energy efficiency of Albanian's buildings stock. *Journal of Cleaner Production*, Vol.38. Part 1, pp.1-15. https://doi.org/10.1016/j.jclepro.2022.135160

[10] Raji Gupta, P., & Singh, R. (2021). PV power forecasting based on data-driven models: a review. In International Journal of Sustainable Engineering, Vol. 14, Issue 6, pp. 1733–1755. *Taylor and Francis Ltd.* https://doi.org/10.1080/19397038.2021.1986590

[11] Mouatasim, A. el, & Darmane, Y. (2018). Regression analysis of a photovoltaic (PV) system in FPO. *AIP Conference Proceedings*, Ouarzazate, Morocco, Vol. 2056, Issue 1 https://doi.org/10.1063/1.5084981

[12] Kallio, S., & Siroux, M. (2023). Photovoltaic power prediction for solar micro-grid optimal control. *Energy Reports*, Vol 9, pp. 594-601. https://doi.org/10.1016/j.egyr.2022.11.081

[13] Efendi, A., Mursityo, Y. T., Hidajati, N. W., Andajani, N., Zuraidah, Z., & Handoyo, S. (2024). Multiple Time Series Modeling of Autoregressive Distributed Lags with Forward Variable Selection for Prediction. *WSEAS Transactions on Business and Economics*, Vol. 21, pp. 1012–1026. https://doi.org/10.37394/23207.2024.21.84

[14] Saroha, S., & Aggarwal, S. K. (2015). A Review and Evaluation of Current Wind Power Prediction Technologies. *WSEAS Transactions on Power Systems*, Vol.10, pp.1-12.

[15] AlShafeey, M., & Csáki, C. (2021). Evaluating neural network and linear regression photovoltaic power forecasting models based on different input methods. *Energy Reports*, Vol. 7, pp. 7601–7614. https://doi.org/10.1016/j.egyr.2021.10.125

[16] Notton, G., Nivet, M. L., Voyant, C., Paoli, C., Darras, C., Motte, F., & Fouilloy, A. (2018). Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting. *Renewable and Sustainable Energy Reviews*, Vol. 87, pp. 96–105. https://doi.org/10.1016/J.RSER.2018.02.007

[17] Keddouda, A., Ihaddadene, R., Boukhari, A., Atia, A., Arıcı, M., Lebbihiat, N., & Ihaddadene, N. (2023). Solar photovoltaic power prediction using artificial neural network and multiple regression considering ambient and operating conditions. *Energy Conversion and Management*, 288, 117186. https://doi.org/10.1016/J.ENCONMAN.2023.117186

[18] IRENA (2020), Global Renewables Outlook: Energy transformation 2050 (Edition: 2020), International Renewable Energy Agency, Abu Dhabi, [Online]. https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2020/Apr/IRENA_Global_Renewables_Outlook_2020.pdf (Accessed Date: October 10, 2024).

[19] Konomi, I., Prifti, V., & Kërpaçi, A. (2022). Evaluating Options to Integrate Energy Storage Systems in Albania. *WSEAS Transactions on Power Systems*, 17, pp.155–168. https://doi.org/10.37394/232016.2022.17.16.

[20] Natural Resources Canada, (2005), Clean Energy Project Analysis Third Edition RETScreen ® Engineering & Cases Textbook, [Online]. https://nrcaniets.blob.core.windows.net/iets/RETScreen_Engineering_e-textbook.pdf (Accessed Date: October 10, 2024).

[21] Natural Resources Canada (2024), RETScreen Expert - Clean Energy Management Software, [Online].. https://natural-resources.canada.ca/maps-tools-and-publications/tools/modelling-tools/retscreen/7465 (Accessed Date: October 10, 2024).

[22] González-Torres, M., Pérez-Lombard, L., Coronel, J. F., Maestre, I. R., & Yan, D. (2022). A review on buildings energy information: Trends, end-uses, fuels and drivers. In *Energy Reports*, Vol. 8, pp. 626–637. Elsevier Ltd. https://doi.org/10.1016/j.egyr.2021.11.280.

[23] ERE, (2023). The Situation of the Power Sector and ERE Activity during 2022. Albanian Energy Regulator Authority (ERE), Tirana, Albania (2023), [Online]. https://www.ere.gov.al/images/files/2023/07/26/Annual_Report_2022.pdf (Accessed Date: October 10, 2024).

[24] Vendim Nr. 537, datë 8.7.2020; Për

miratimin e kërkesave minimale të performancës së energjisë së ndërtesave dhe të elementeve të ndërtesave, [Online]. https://eficenca.gov.al/wp-content/uploads/2022/09/Vendim-Nr.-537-date-8.7.2020-per-kerkesat-minimale-te-performances-se-energjise.pdf (Accessed Date: October 10, 2024).

[25] Law no. 116/2016. "On Building Energy Performance", [Online]. https://eficenca.gov.al/wp-content/uploads/2022/09/Ligj-nr.-1162016-per-performancen-e-energjise-se-ndertesave.pdf (Accessed Date: October 10, 2024).

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

- Raimonda Dervishi: Conceptualization of the published work, formulation, and evolution of overarching research goals and aims. Data curation and scrubbing data and maintaining research data including proofing and validation.
- Erjola Cenaj: Formal analysis and preparation, creation, and presentation of the published work. Data curation and scrubbing data and maintaining research data including proofing and validation.
- Lorenc Malka: Conceptualization of the published work, formulation, and evolution of overarching research goals and aims. Formal software analysis and simulations in RETScreen Expert model.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare.