# Various multistage ensembles for prediction of heating energy consumption

R. Jovanović [1]  A. Sretenović [2]

[1] *Department of Automatic Control, Faculty of Mechanical Engineering University of Belgrade, Kraljice Marije 16, Belgrade, Serbia. E-mail: rjovanovic@mas.bg.ac.rs*

[2] *Department of Thermal Science, Faculty of Mechanical Engineering University of Belgrade, Kraljice Marije 16, Belgrade, Serbia. E-mail: asretenovic@mas.bg.ac.rs*

## Abstract

Feedforward neural network models are created for prediction of daily heating energy consumption of a NTNU university campus Gløshaugen using actual measured data for training and testing. Improvement of prediction accuracy is proposed by using neural network ensemble. Previously trained feed-forward neural networks are first separated into clusters, using k-means algorithm, and then the best network of each cluster is chosen as member of an ensemble. Two conventional averaging methods for obtaining ensemble output are applied; simple and weighted. In order to achieve better prediction results, multistage ensemble is investigated. As second level, adaptive neuro-fuzzy inference system with various clustering and membership functions are used to aggregate the selected ensemble members. Feedforward neural network in second stage is also analyzed. It is shown that using ensemble of neural networks can predict heating energy consumption with better accuracy than the best trained single neural network, while the best results are achieved with multistage ensemble.

*Keywords:* heating consumption prediction, multistage neural network ensemble, adaptive neuro-fuzzy inference

## 1 Introduction

The study of the building energy demand has become a topic of great importance, because of the significant increase of interest in energy sustainability, especially after the emanation of the EPB European Directive. In Europe, buildings account for 40% of total energy use and 36% of total $CO_2$ emission Council. (2010). According to Bergesen et al. (2013), 66% of the total energy consumption of residential buildings occurs in the space heating sector of Norwegian residential buildings. Therefore, the estimation or prediction of building energy consumption has played very important role in building energy management, since it can help to indicate above-normal energy use and/or diagnose the possible causes, if there has been enough historical data gathered. Scientists and engineers are lately moving from calculating energy consumption toward analyzing the real energy use of buildings. One of the reasons is that, due to the complexity of the building energy systems and behavior, non-calibrated models cannot predict well building energy consumption, so there is a need for real time image of energy use (using measured and analyzed data).

The classic approach to estimate the building energy use is based on the application of a model with known system structure and properties as well as forcing variables (forward approach). Using different software tools, such as EnergyPlus, TRNSYS, BLAST, ESP-r, HAP, APACHE requires detailed knowledge of the numerous building parameters (constructions, systems) and behavior, which are usually not available.

Perera et al. (2014), developed continuous time mathematical heating model for a building unit based on the first principles. The developed model was implemented in a MATLAB environment, and mainly a theoretical approach is used to validate it for a residential building unit. Model is also validated using experimental data.

In recent years, considerable attention has been given to a different approach for building energy analysis, which is based on the so called "inverse" or data-driven models Kusiak et al. (2010). In a data-driven approach, it is required that the input and output variables are known and measured, and the development of the "inverse" model consists in determination of a mathematical description of the relationship between the independent variables and the dependent one. The data-driven approach is useful when the building (or a system) is already built, and actual consumption (or performance) data are measured and available. For this approach, different statistical methods can be used. Artificial neural networks (ANN) are the most used artificial intelligence models for different types of prediction. The main advantages of an ANN model are its self-learning capability and the fact that it can approximate a nonlinear relationship between the input variables and the output of a complicated system. Feedforward neural networks are most widely used in energy consumption prediction. Ekici and Aksoy (2009) proposed a backpropagation three-layered ANN for the prediction of the heating energy requirements of different building samples. Dombaycı (2010) used hourly heating energy consumption for a model house calculated by degree-hour method for training and testing the ANN model. In Ekonomou (2010) actual recorded input and output data that influence Greek long-term energy consumption were used in the training, validation and testing process. Li et al. (2011) proposed the hybrid genetic algorithm-adaptive network-based fuzzy inference system (ANFIS) which combined the fuzzy if-then rules into the neural network-like structure for the prediction of energy consumption in the library building. The calculated results indicated better performance compared with ANN in terms of forecasting accuracy. An excellent review of the different neural network models used for building energy use prediction was done by Kumar et al. (2013). The ensemble of neural networks is a very successful technique where the outputs of a set of separately trained neural networks are combined to form one unified prediction, Zhou et al. (2002). Since an ensemble is often more accurate than its members, such a paradigm has become a hot topic in recent years and has already been successfully applied to time series prediction Melin et al. (2012), weather forecasting Taylor and Buizza (2002), load prediction in a power system Siwek et al. (2009).

The main idea of this paper is to propose multistage neural network ensemble for prediction of heating energy use. The ensemble members are chosen among 50 separately trained feedforward neural networks using k-means clustering, and for combining their outputs different ANFIS in second stage are used.

# 2 Feedforward neural network (FFNN)

Artificial neural network (ANN) method is a computational intelligence technique, based on the information processing system of the human brain and which may be used as an alternative method in engineering analysis and prediction. ANNs work as a black-box model, thus, it is not necessary to have detailed information about the system. Instead, they learn the relationship between input and output variables by means of historical data, similar to the way a nonlinear regression might perform Karatasou et al. (2006). The FFNN architecture consists of an input layer, an output layer, and one or more hidden layers of interconnected processors called neurons. Each layer has a number of neurons and each neuron is fully interconnected with adaptable weighted connections to neurons in the subsequent layer. Therefore, each neuron receives input signals from other neurons or external stimuli, processes it locally through an activation function and produces a transformed output signal to other neurons or external outputs. The nonlinear activation functions in the hidden layer neurons enable the neural network to be a universal approximator. The process of training network is the adjustment of the weights, so that the network can produce the desired response to the given inputs. Different training algorithms could be applied to minimize the error function, but the most widely used are the backpropagation algorithm and the algorithms derived from it. They use a gradient descent technique to minimize the cost function which is the mean square difference between the desired and the actual network outputs. In this study, a multilayer feedforward network with single hidden layer and backpropagation learning algorithm (BPNN) is used. In BPNN the learning algorithm has two phases. First a training input data set is presented to the network input layer. The network then propagates the input data set from layer to layer until the output data set is generated by the output layer. If this data set is different from the desired output, an error is calculated and then propagated backwards through the network from the output layer to the input layer. The weights are modified as the error is propagated.

# 3 Artificial neural network ensembles

Many engineering problems, especially in energy use prediction, appeared to be too complex for a single neural network. Researchers have shown that simply combining the output of many neural networks can generate more accurate predictions and significantly improve generalization ability than that of any of the individual networks Hansen and Salamon (1990). Theoretical and empirical work showed that a good ensemble is one where the individual networks have both accuracy and diversity, namely the individual networks make their errors on different parts of the input space Qiang et al. (2005). An important problem is, then, how to select the aggregate members in order to have an optimal compromise between these two conflicting conditions Granitto et al. (2005). The accuracy can be described by the mean square error (or some other prediction indicator) and achieved by proper training algorithms of neural networks. Diverse individual predictors (members) can be obtained in several ways. The most widely used approaches Sharkey (1999), Zhang et al. (2001) can be divided in three groups. The first group of methods refers to training individuals on different adequately-chosen subsets of the data set. It includes elaborations of two important techniques: bagging and boosting. Bagging is proposed by Breiman (1996) based on bootstrap sampling. It creates several training sets from the original training set and then trains a component neural network from each of those training sets. Boosting, proposed by Schapire (1990), generates a series of component neural networks whose training sets are determined by the performance of former ones. Training instances that are wrongly predicted by former networks will play more important roles in the training of later networks. The second group uses variation of topologies, by varying number of input and/or hidden nodes, initial weight sets, training algorithms, or even networks with different types. The third group is named selective approach group where the diverse components are selected from a number of accurately trained networks. Opitz and Shavlik (1996) proposed a generic algorithm to search for a highly diverse set of accurate networks. Other used algorithms for selecting ensemble components are: pruning algorithm to eliminate redundant classifiers Lazarevic and Obradovic (2001), selective algorithm based on bias/variance decomposition Navone et al. (2000), genetic algorithm (GASEN) proposed by Zhou et al. (2001) and PSO based approach proposed by Fu et al. (2004). Clustering technology can be used to divide all networks into some groups (clusters) according to similarity of the networks. Then, one most accurate individual in each group on the validation set is selected, and finally, all selected individuals construct the ensemble.

# 4 K-means for selecting ensemble members

In Qiang et al. (2005) the clustering-based selective neural network (using k-means) was compared with two main ensemble approaches: Bagging and Boosting. K-means clustering, proposed by MacQueen et al. (1967) is a method commonly used to automatically partition a data set into $m$ groups. Even though k-means was first proposed over 50 years ago, it is still one of the most widely used algorithms for clustering. Ease of implementation, simplicity, efficiency, and empirical success are the main reasons for its popularity. This technique is based on distance matrix, using Euclidean distance as a criterion. It starts with $m$ initial cluster centers and for all data, Euclidean distance from each cluster center is calculated, after which the data points are assigned to the closest cluster center. This method is being repeated until the squared error between the empirical mean of a cluster and the points in the cluster is minimized. When using k-means for selecting neural network ensemble members, the goal is to divide prediction data achieved by individual networks $y = \{y_1, \ldots, y_r\}$ into $m$ clusters, where number of elements in each cluster is $n_i$, and the center of cluster is $c_i$. So clustering can be achieved by finding $c_i$ which makes

$$J_e = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \|y_j^{(i)} - c_i\|^2 \qquad (1)$$

minimized. Obviously, after clustering the diversity between networks in different cluster groups is greater than those within the same group. The diversity is maintained by choosing the most accurate networks in each group as a member of the ensemble. In k-means algorithm, cluster number $m$ must be determined in advance. To select the best $m$ value, the prediction indices of the created ensemble can be compared. Linear combination of the outputs of ensemble members is one of the most popular approaches for combining selected network outputs (simple or weighted). Different approach comprises using neural network for combining selected ensemble members. Ilić et al. (2012) proposed the system comprised of two ANNs assembled in a hierarchical order. In this paper, for the second stage, adaptive neuro-fuzzy inference system (ANFIS) is proposed.

# 5 Adaptive Neuro-Fuzzy Inference System (ANFIS)

The process of fuzzy inference involves membership functions, fuzzy logic operators, and if-then rules. Fuzzy inference systems (FIS) have been successfully applied in fields such as automatic control Skullestad et al. (2001), monitoring and maintenance Cibulka et al. (2012), data classification, decision analysis, expert systems, and computer vision. Overview of possible application of fuzzy logic in modeling, identification and control can be found in Zadeh (1994). The adaptive network-based fuzzy inference system (ANFIS) proposed by R. Jang Jang (1993) is one of the most commonly used fuzzy inference systems, and its architecture is obtained by embedding FIS into the framework of adaptive networks. The generalization capability of the fuzzy logic is very poor because it uses the heuristic algorithms for defuzzification, rule evolution and antecedent processing. The main disadvantage of neural network is how to determine proper size and optimal structure of the network. Also, the relationships of weight changes with input output behavior during the training and use of trained system to generate correct output using the weights are very complicated to understand, like a "black box". Combining fuzzy logic and neural network is preeminent idea to overcome the disadvantages of both techniques. Neural networks are used to tune the membership functions of fuzzy systems even for complex systems, Singh et al. (2012). The outstanding property of ANFIS is that it compensates the disadvantage of FIS with the learning mechanism of NN. The architecture of the ANFIS used in this study is based on the first-order Takagi-Sugeno model Takagi and Sugeno (1985). For a simple MISO system (multi-input, single-output), having two inputs $(x_1, x_2)$ and one output $(y)$, typical rule set can be expressed as:

$$\textbf{Rule1}: \textbf{If } x_1 \textbf{ is } A_1 \textbf{ and } x_2 \textbf{ is } B_1 \qquad (2)$$
$$\textbf{then } f_1 = a_1 x_1 + b_1 x_2 + c_1$$
$$\textbf{Rule2}: \textbf{If } x_1 \textbf{ is } A_2 \textbf{ and } x_2 \textbf{ is } B_2$$
$$\textbf{then } f_2 = a_2 x_1 + b_2 x_2 + c_2.$$

The ANFIS architecture is shown in Figure 1. It is composed of five layers where each layer contains several nodes described by the node function. Let $O_i^j$ denote the output of the $i$-th node in layer $j$.

**Layer 1**: In the first layer, all the nodes are adaptive nodes. The outputs of layer 1 are the fuzzy membership grades of the inputs, which are given by:

$$O_i^1 = \mu_{A_i}(x_1), \quad i = 1, 2, \qquad (3)$$
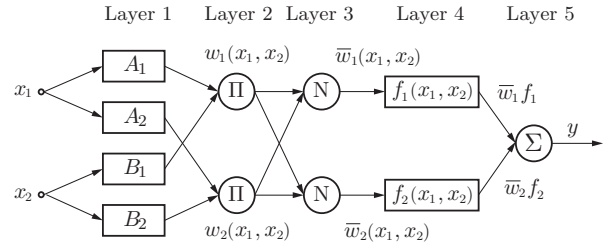$$O_i^1 = \mu_{B_{i-2}}(x_2), \quad i = 3, 4.$$



Figure 1: ANFIS network architecture

where $A_i$ and $B_i$ are the linguistic labels and $\mu_{A_i}$ and $\mu_{B_i}$ are the membership functions for $A_i$ and $B_i$ linguistic labels, respectively. As node functions in this layer any continuous and piecewise differentiable functions, such as commonly used trapezoidal, triangular-shaped, Gaussian or generalized bell membership functions, can be used. Therefore, outputs of this layer form the membership values of the premise part and parameters contained in membership functions of fuzzy sets called premise parameters.

**Layer 2**: In contrast to layer 1 the nodes in this layer are fixed. The output $O_i^2$ of the node $i$ can be computed as:

$$O_i^2 = w_i = \mu_{A_i}(x_1) \cdot \mu_{B_i}(x_2), \quad i = 1, 2. \qquad (4)$$

where $w_i$ represents a firing strength of a rule.

**Layer 3**: In this layer where the normalization process is performed, the nodes are fixed. The ratio of the $i$-th rules firing strength to the sum of all rules firing strengths is calculated for the corresponding node and thus the outputs of this layer are called normalized firing strengths:

$$O_i^3 = \overline{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2. \qquad (5)$$

**Layer 4**: The fourth layer deals with the consequent part of the fuzzy rule. Every node $i$ in this layer is an adaptive node and it calculates the contribution of $i$-th rule in the model output function which is defined based on the first-order Takagi-Sugeno method as:

$$O_i^4 = \overline{w}_i f_i = \overline{w}_i (a_i x_1 + b_i x_2 + c_i), \quad i = 1, 2. \qquad (6)$$

where $\{a_i, b_i, c_i\}$ is the parameter set. Parameters in this layer are referred to as consequent parameters.

**Layer 5**: This is the summation layer, which consists of a single fixed node. It sums up all the incoming signals and produces the output:

$$O_i^5 = y = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}. \qquad (7)$$

From the proposed ANFIS architecture, it is observed that given the values of premise parameters,

the overall output can be expressed as a linear combinations of the consequent parameters. More precisely, the output $y$ can be rewritten as:

$$y = \frac{w_1}{w_1 + w_2}f_1 + \frac{w_2}{w_1 + w_2}f_2 = \overline{w}_1 f_1 + \overline{w}_2 f_2 \quad (8)$$
$$= (\overline{w}_1 x_1)a_1 + (\overline{w_1 x_2})b_1 + \overline{w}_1 c_1$$
$$+ (\overline{w}_2 x_1)a_2 + (\overline{w}_2 x_2)b_2 + \overline{w}_1 c_2.$$

In the training process, the least squares method (forward pass) is used to optimize the consequent parameters with the premise parameters fixed. Once the optimal consequent parameters are found, the backward pass starts immediately. The gradient descent method (backward pass) is used to adjust optimally the premise parameters corresponding to the fuzzy sets in the input domain. The output of the ANFIS is calculated by employing the consequent parameters found in the forward pass. The output error is used to adapt the premise parameters by means of a standard back propagation algorithm.

# 6 Case study

According to Bergesen et al. (2013), 66% of the total energy consumption of residential buildings occurs in the space heating sector of Norwegian residential buildings. University campuses are specific groups of diverse buildings, with significant energy consumption, Sretenovic (2013). They consist of many different buildings, representing small-scale town for itself. Therefore, they provide an excellent testbed to characterize and understand energy consumption of group of mixed use buildings. Norwegian University of Science and Technology (NTNU) campus Gløshaugen consists of 35 buildings, with total area of approximately 300,000 $m^2$. Building and Energy Management System (BEMS) and web-based Energy Monitoring System (Energy Remote Monitoring-ERM) are available at NTNU. The Schneider ERM system is an Automatic Monitoring and Targeting system with advanced analysis features, which receives main meter and sub-meter consumption data and provides system energy reporting, alarming, monitoring and analysis. There are 46 heating meters installed in campus. Hourly heat and electricity consumption from all meters can be collected on ERM (EnergyRemoteMonitoring (2014)). District heating net is organized in form of the ring, while the main heat exchanger is installed in Old electric building (Figure 2). The Main meter is installed by the district heating supplier, so it was taken as relevant. Daily heating energy consumption was analyzed in this paper. Creating a model of energy use helps in future building planning; it can provide useful information about most probable energy consumption for

similar buildings, or predict energy use in different conditions. Also these models can be used to show impacts of possible energy savings measures and help in finding optimal way of reducing energy costs. It is also very important to have correct and reliable measured data. If a part of a building is leased to other users (which is the case in campus Gløshaugen), there is necessity for calculating bills for each tenant. There is increased interest in data error analysis and developing methods that can point out possible meters malfunction. Also, without correct measured data it is not possible to monitor and prove benefits of applying energy saving measures for increasing energy efficiency. Creating representative model of heating energy consumption can also indicate errors in measured data.
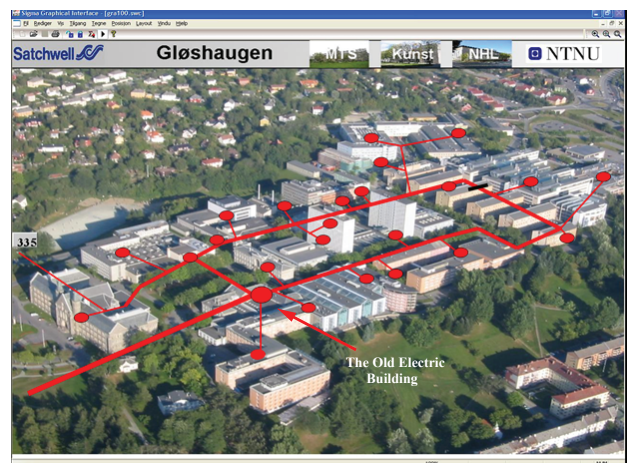


Figure 2: University campus Gløshaugen

## 6.1 Data pre-processing

All weather data were gathered from the local meteorological station Skjetlein AgroMetbase (2014). Heating season in Trondheim area lasts around 251 to 280 days, Tveito (2002). The heating season is defined as the period from the day the mean daily temperature falls below 11°C during the autumn and until the day it rises above 9°C during the spring Johannessen (1956). Based on the analysis for the period 1961-1990, the beginning of the heating season in Trondheim is usually between 29/08-17/09, and the end of the heating season 10/05-29/05 Tveito (2002). Considering that the outside temperature has the biggest influence on heating energy consumption, mean daily outside temperatures for years 2006 until 2014 were investigated in order to determine optimal number of neural networks. The average mean daily temperatures for the last 8 years is shown in Figure 4. The bars show the maximum and minimum mean daily temperature for
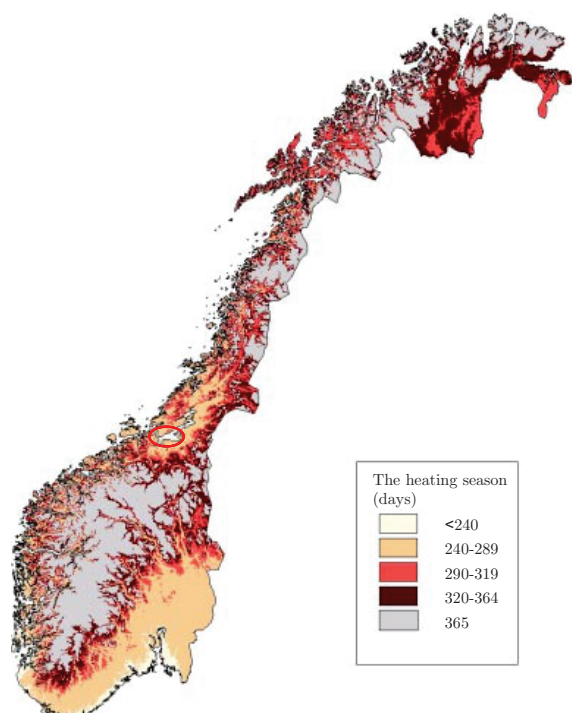
Figure 3: Heating season in Norway Tveito (2002)



Figure 5: Correlation of the daily heating energy consumption with mean daily outside temperature for the year 2012

the specific date in the same period (years 2006-2014). After the analysis, database is divided as follows:

- Cold period  from January $1^{st}$ until March $31^{st}$ and from November $1^{st}$ until December $31^{st}$

- Mild period  From April $1^{st}$ until June $15^{th}$ and from September $16^{th}$ until October $31^{st}$

- Warm period (outside of heating season) - June $15^{th}$ until September $15^{th}$ is excluded from the analysis

It implicates that better prediction results can be obtained using separate network models for each period compared to using one network for all year. In this paper, only the cold period (with biggest heating energy consumption) is analyzed.

The daily heating consumption is analyzed in terms of the type of the day. The correlation with mean daily outside temperature for each day of the week for the year 2012 is shown in Figure 5. Analysis showed that there is no specific difference between the working days (heating consumptions for Monday to Friday have similar trendlines), while the regression lines for Saturday and Sunday are below them, as expected. In NTNU campus Gløshaugen, heating is not switched off during the weekends, only the design set-point is lowered, so the heating consumption on Monday is not significantly different than the other working days. The analysis of
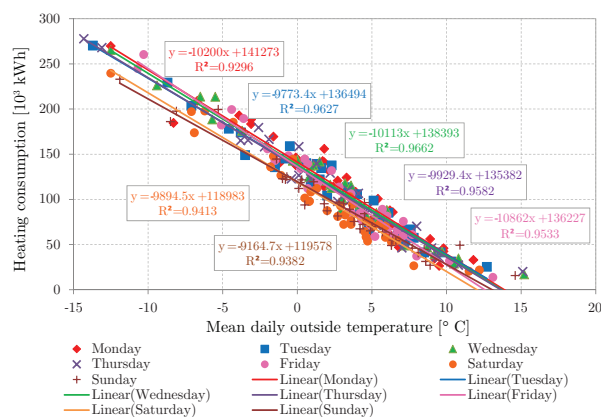
the daily heating consumption also showed that during the holidays and exam periods, heating operation is at the same level as for the working days (heating is maintained at the designed set-point). These conclusions implicate that there should be two separate networks created: one for the working days, and other for the weekend. In this paper, the network for the working days is analyzed.

# 7 ANN models development

The most important task in building an ANN prediction model is the selection of input variables. Many different studies dealing with impact of various variables on energy consumption can be found in literature. Empirical research of the influence of hourly values of solar radiation and wind speed on heating demands of building complex heated by district heating system was conducted in Wojdyga (2008). The research results confirmed the influence of increasing heat demand in case of higher wind speeds and decreasing heat demand in cases of sunny days occurring during the heating season. All input variables for the neural network model, that are considered in this study, are: mean daily outside temperature [°C], mean daily wind speed [$m/s$], total daily solar radiation [$Wh/m^2$], minimum daily temperature [°C], maximum daily temperature [°C], relative humidity [%], day of the week, month of the year and heating consumption of the previous day. Partial autocorrelation, which measures how a series is correlated with itself at different lags, indicate that the heating consumption of the previous day has the biggest influence on the heating consumption of the observed day (Figure 6). Therefore, the heating consumption of the previous day is selected as additional
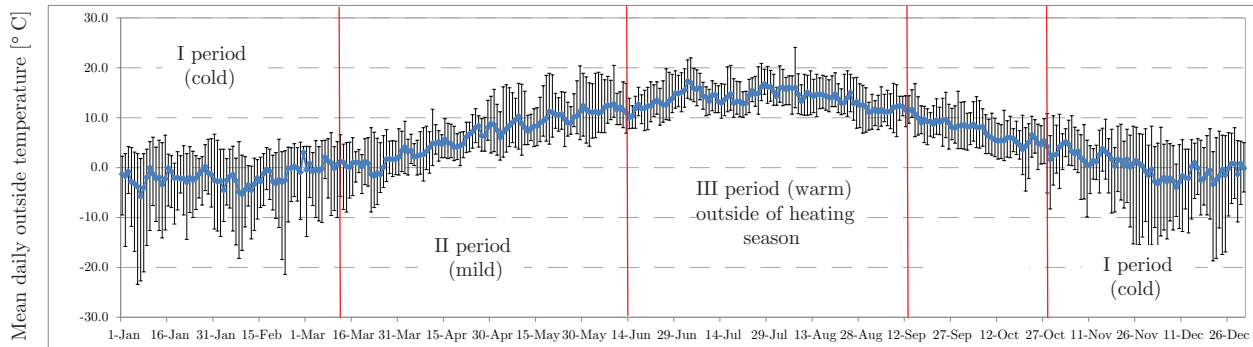
Figure 4: Average mean daily temperature for years 2006-2014

input variable. In that way, the prediction is always done for one day ahead. For long-term forecast it is necessary to use this model to perform prediction day by day. However, in that case, the prediction error is accumulated. But, even in the case of static model, where the values of inputs and/or outputs variables of the model for the previous day are not used, for the prediction for longer period in advance it is necessary to have input variables (temperature, wind speed, etc.) for that period. One way is to develop models to separately predict these input variables and then use them to predict consumption, which would again result in error being accumulated at the end.
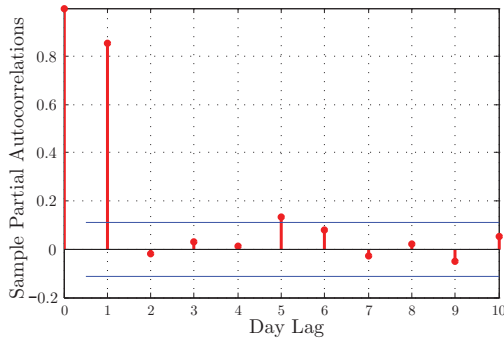


Figure 6: Partial autocorrelation function

The ANN architecture used in this study is a three-layer FFNN composed of one input layer, one output layer and one hidden layer, with LevenbergMarquardt learning algorithm. In the hidden layer and output layer tansig (sigmoidal) and linear (purelin) activation functions are used. During the application study many different values of hidden neurons were examined using trial and error method, and the best results are achieved with one hidden layer with 10 neurons. For

training the models, data for the working days in the cold period (from January $1^{st}$ until March $31^{st}$ and from November $1^{st}$ until December $31^{st}$) for years 2009, 2010 and 2011 were used (318 samples in total), and for testing 2012 (100 samples). Data with obvious errors and heat meter malfunctions were removed from the dataset. To ensure that no special factor is dominant over the others, all inputs and outputs are normalized to the interval (0, 1) by a linear scaling function. The prediction accuracy of all proposed models is measured by the coefficient of determination ($R^2$), root mean square error (RMSE) and mean absolute percentage error (MAPE).

## 7.1 Neural network ensemble

Possible improvement of the prediction accuracy by using network ensemble is examined. The application of an ensemble technique is divided into two steps. As the first step, after training numerous FFNNs, 50 networks with satisfying accuracy are selected for possible members. The second step is the adequate combination of outputs of the ensemble members to produce the most appropriate output. In order to improve ensemble efficiency, we need to ensure both accuracy of networks and diversity between individuals. The diversity is achieved by appropriate selection of members from many previously trained networks. Considering the difficulty of selecting diversity and accuracy at the same time, we can apply an easier method to gradually achieve both goals. First, we employ clustering technology to divide all networks into some groups (clusters) according to similarity of the networks. Then, one most accurate individual in each group on the validation set is selected. Finally, all selected individuals construct the ensemble, as it can be seen in Figure 7.

There are different methods for combining the outputs. The conventional approach is to use averaging:
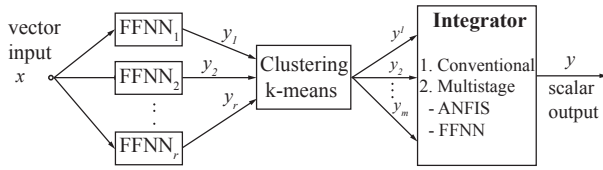
125

Figure 7: Proposed neural network ensemble

- ANFIS trimf
- ANFIS gbellmf
- ANFIS gaussmf
- ANFIS FCM
- ANFIS SUB

3. FFNN multistage

simple or weighted. The other method, multistage approach, which is expected to give an even better improvement in accuracy, would be to use another neural network as an integrator of the individual classifiers. Two different network architectures are proposed in second stage: FFNN and ANFIS. Generally, structure identification in fuzzy modelling involves several tasks: selecting input variables, input space partitioning, choosing the number and types of membership function for inputs, creating fuzzy rules an selecting initial parameters for membership function. In this paper, different ANFIS models are constructed using three different identification methods: grid partitioning, subtractive clustering and fuzzy C-means clustering. In the grid partitioning method, the domain of each antecedent variable is partitioned into equidistant and identically shaped membership functions, which are previously defined. To demonstrate the effect of choice of MF on the model performance three different functions are tested: the triangular MF (trimf), the generalized bell MF (gbellmf) and the Gaussian MF (gaussmf). The number of MFs for each of the input of ANFIS is set to 2. Fuzzy C-means (FCM), developed by Dunn (1973) and improved by Bezdek (1981), clusters the data by minimizing the total distance of each data point to the cluster centers, wherein each data point belongs to a cluster to some degree that is specified by a membership grade. Subtractive clustering is one of the automated data-driven based methods for constructing the primary fuzzy models proposed by Chiu (1994). It is a fast, one-pass algorithm for estimating the number of clusters and the cluster centers in a set of data, based on the density of data points in input space. As a results, fuzzy model with minimum number of rules is obtained. In the second stage, FFNN with the same architecture as in the first stage is used. All proposed ensembles that are analyzed and compared with the best trained single FFNN:

1. Conventional ensemble:
   - simple average
   - weighted average

2. ANFIS multistage:

# 8 Results and discussion

Most critical task in selection of user-specified parameters required by k-means algorithm is the choice of number of clusters. Since there is no perfect mathematical criterion, numerous heuristics and index methods are available in literature. The other way is to run k-means independently for different cluster numbers, and select the one where ensemble prediction achieves best results. The grid partitioning is only suitable for cases with small number of input variables, because the number of fuzzy rules increases exponentially when the number of fuzzy rules increases (curse of dimensionality problem). In this study, number of input variables for second stage neural network is defined by the number of clusters used in k-means clustering. Due to the mentioned curse of dimensionality problem, number of clusters is varied from 2 to 6, so different multistage ANFIS methods can be equally compared.

The number of networks in the ensemble is equal to cluster number because one best network in each cluster is selected to join the ensemble. In Table 1 and Table 2 prediction indices for training and testing the models, respectively, are presented.

The presented results show that all neural network models can predict heating consumption with satisfying accuracy. Even the single best trained FFNN gives satisfactory values of 0.9773 for $R^2$ and 6.3049% for MAPE in testing period. Both conventional methods for creating ensemble show improvement in accuracy, with weighted average for 6 clusters (ensemble members) $R^2$ raised to 0.9818, while MAPE is 5.6270%. Further enhancement can be achieved by second layer network used to combine the outputs of the individual ensemble members, both FFNN and ANFIS. The main idea of this paper is not to specify the best second stage NN architecture, optimal membership function, or ANFIS clustering type, but to show the general ability of various multistage ensembles to successfully predict heating consumption. The best result for $R^2$ is 0.9829, achieved with multistage FFNN with 6 cluster, and lowest MAPE is 5.3383%, using ANFIS-FCM multistage. In all models, ensemble has proven its preeminence comparing to the best trained single FFNN.

Table 1: Prediction indices for training networks

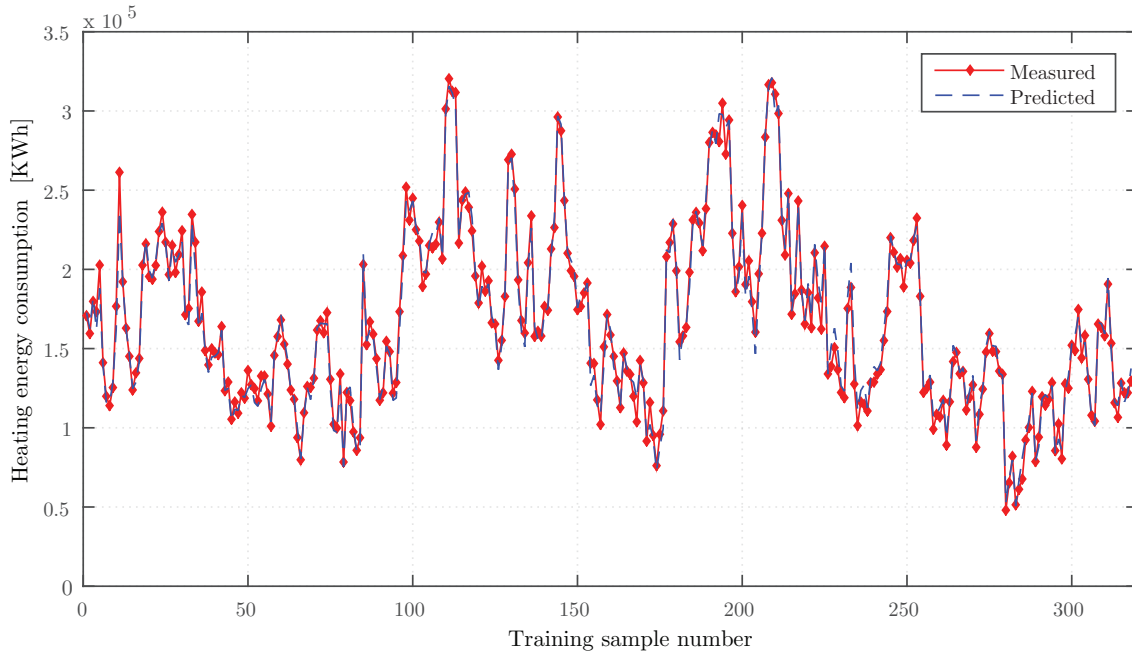| Number of clusters | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| **$R^2$ [−]** | | | | | |
| Best FFNN | 0.9850 | 0.9850 | 0.9850 | 0.9850 | 0.9850 |
| Ensemble SAV | 0.9866 | 0.9863 | 0.9863 | 0.9866 | 0.9866 |
| Ensemble WAV | 0.9866 | 0.9863 | 0.9865 | 0.9870 | 0.9870 |
| MS ANFIS-trimf | 0.9869 | 0.9878 | 0.9883 | 0.9893 | 0.9899 |
| MS ANFIS-gbellmf | 0.9873 | 0.9874 | 0.9882 | 0.9897 | 0.9907 |
| MS ANFIS-gaussmf | 0.9872 | 0.9875 | 0.9880 | 0.9899 | 0.9904 |
| MS ANFIS-FCM | 0.9868 | 0.9869 | 0.9869 | 0.9880 | 0.9883 |
| MS ANFIS-SUB | 0.9871 | 0.9872 | 0.9867 | 0.9875 | 0.9875 |
| MS ANFIS-FFNN | 0.9870 | 0.9864 | 0.9867 | 0.9875 | 0.9874 |
| **RMSE [kWh]** | | | | | |
| Best FFNN | 6942.4 | 6942.4 | 6942.4 | 6942.4 | 6942.4 |
| Ensemble SAV | 6445.8 | 6648.0 | 6524.1 | 6417.6 | 6407.0 |
| Ensemble WAV | 6447.2 | 6550.4 | 6440.4 | 6312.2 | 6294.6 |
| MS ANFIS-trimf | 6309.7 | 6097.1 | 5974.1 | 5710.8 | 5543.0 |
| MS ANFIS-gbellmf | 6216.2 | 6188.3 | 5999.3 | 5584.7 | 5314.8 |
| MS ANFIS-gaussmf | 6233.9 | 6167.6 | 6038.1 | 5550.6 | 5393.8 |
| MS ANFIS-FCM | 6331.7 | 6322.3 | 6310.3 | 6029.4 | 5961.2 |
| MS ANFIS-SUB | 6275.7 | 6246.1 | 6352.4 | 6174.2 | 6163.2 |
| MS ANFIS-FFNN | 6378.9 | 6508.1 | 6357.5 | 6173.5 | 6273.3 |
| **MAPE [%]** | | | | | |
| Best FFNN | 3.5614 | 3.5614 | 3.5614 | 3.5614 | 3.5614 |
| Ensemble SAV | 3.3706 | 3.3524 | 3.2418 | 3.2276 | 3.2403 |
| Ensemble WAV | 3.3654 | 3.3193 | 3.2169 | 3.2278 | 3.2329 |
| MS ANFIS-trimf | 3.2178 | 3.0928 | 3.0103 | 2.8664 | 2.8095 |
| MS ANFIS-gbellmf | 3.1658 | 3.1239 | 3.0692 | 2.8454 | 2.7584 |
| MS ANFIS-gaussmf | 3.1765 | 3.1088 | 3.0536 | 2.8347 | 2.7766 |
| MS ANFIS-FCM | 3.3056 | 3.1718 | 3.1570 | 2.9960 | 2.9922 |
| MS ANFIS-SUB | 3.2882 | 3.1196 | 3.1711 | 3.0835 | 3.0892 |
| MS ANFIS-FFNN | 3.2621 | 3.2684 | 3.1456 | 3.1479 | 3.2779 |



Figure 8: Prediction results of multistage ensemble ANFIS FCM with 5 clusters for training period

Table 2: Prediction indices for testing networks

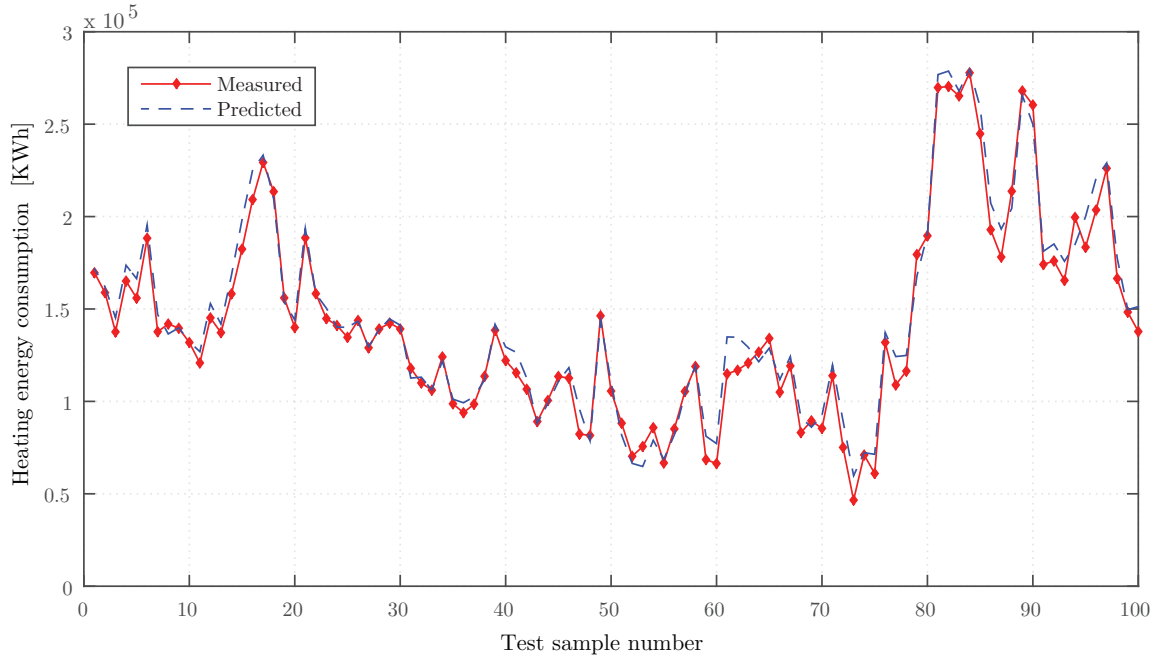| Number of clusters | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| **R$^2$ [−]** | | | | | |
| Best FFNN | 0.9773 | 0.9773 | 0.9773 | 0.9773 | 0.9773 |
| Ensemble SAV | 0.9806 | 0.9804 | 0.9803 | 0.9811 | 0.9816 |
| Ensemble WAV | 0.9806 | 0.9801 | 0.9801 | 0.9814 | 0.9818 |
| MS ANFIS-trimf | 0.9802 | 0.9803 | 0.9804 | 0.9803 | 0.9803 |
| MS ANFIS-gbellmf | 0.9809 | 0.9814 | 0.9808 | 0.9796 | 0.9792 |
| MS ANFIS-gaussmf | 0.9803 | 0.9814 | 0.9808 | 0.9806 | 0.9814 |
| MS ANFIS-FCM | 0.9810 | 0.9813 | 0.9802 | 0.9814 | 0.9811 |
| MS ANFIS-SUB | 0.9815 | 0.9813 | 0.9801 | 0.9818 | 0.9818 |
| MS ANFIS-FFNN | 0.9807 | 0.9811 | 0.9807 | 0.9810 | 0.9829 |
| **RMSE [kWh]** | | | | | |
| Best FFNN | 9829.7 | 9829.7 | 9829.7 | 9829.7 | 9829.7 |
| Ensemble SAV | 9101.2 | 9466.1 | 9044.5 | 8877.1 | 8710.5 |
| Ensemble WAV | 9065.5 | 9228.0 | 8902.9 | 8720.5 | 8644.3 |
| MS ANFIS-trimf | 8556.0 | 8757.7 | 8658.3 | 8510.9 | 8510.6 |
| MS ANFIS-gbellmf | 8405.1 | 8460.1 | 8666.9 | 8604.8 | 8583.9 |
| MS ANFIS-gaussmf | 8476.3 | 8572.5 | 8564.6 | 8608.8 | 8334.5 |
| MS ANFIS-FCM | 8674.8 | 8650.7 | 8618.7 | 8221.2 | 8237.6 |
| MS ANFIS-SUB | 8584.1 | 8609.2 | 8715.2 | 8402.6 | 8375.3 |
| MS ANFIS-FFNN | 8639.6 | 9101.1 | 8884.2 | 8544.3 | 8813.3 |
| **MAPE [%]** | | | | | |
| Best FFNN | 6.3049 | 6.3049 | 6.3049 | 6.3049 | 6.3049 |
| Ensemble SAV | 5.8460 | 6.2800 | 6.0255 | 5.8500 | 5.7336 |
| Ensemble WAV | 5.8198 | 6.0304 | 5.9122 | 5.6944 | 5.6270 |
| MS ANFIS-trimf | 5.5213 | 5.6406 | 5.7413 | 5.6930 | 5.6309 |
| MS ANFIS-gbellmf | 5.3742 | 5.4347 | 5.6460 | 5.5810 | 5.6839 |
| MS ANFIS-gaussmf | 5.4119 | 5.5030 | 5.5998 | 5.5332 | 5.3887 |
| MS ANFIS-FCM | 5.5972 | 5.5766 | 5.6459 | 5.3810 | 5.3883 |
| MS ANFIS-SUB | 5.5521 | 5.4382 | 5.7044 | 5.4872 | 5.5686 |
| MS ANFIS-FFNN | 5.6625 | 5.8751 | 5.7089 | 5.6053 | 5.7987 |



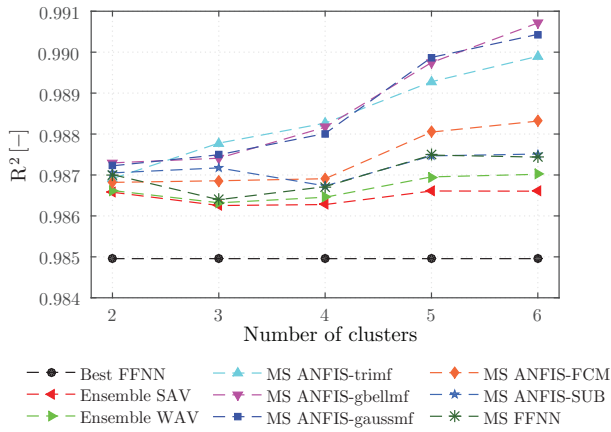Figure 13: Prediction results of multistage ensemble ANFIS FCM with 5 clusters for testing period
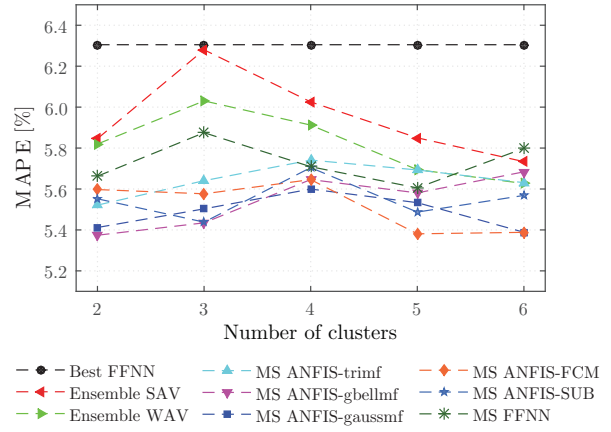
Figure 9: R$^2$ for different cluster number for training
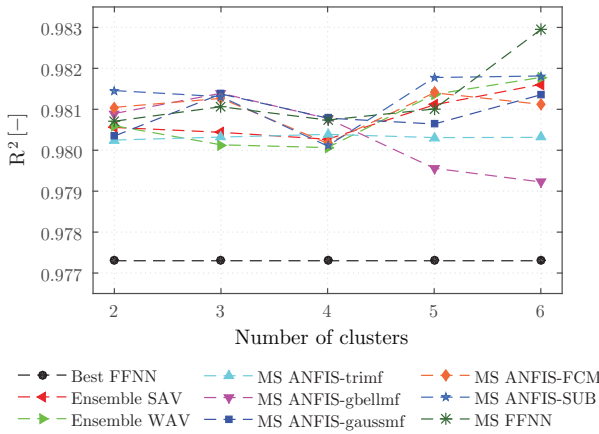


Figure 10: R$^2$ for different cluster number for testing
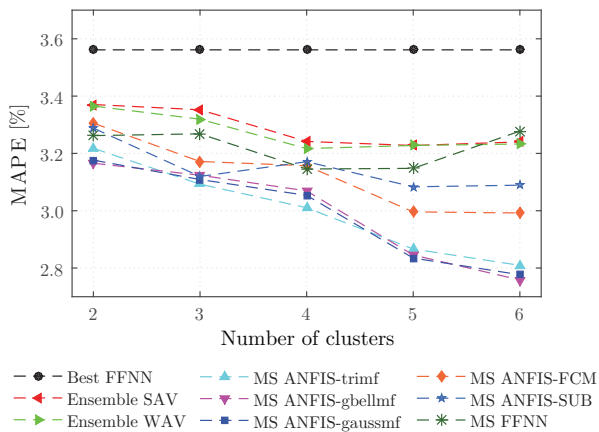


Figure 11: MAPE for different cluster number for training



Figure 12: MAPE for different cluster number for testing

members can be seen in Figure 9 (training) and Figure 10 (testing). MAPE for different cluster numbers for training and testing period are shown in Figure 11 and Figure 12, respectively. Figure 8 and Figure 13 present the comparison of the prediction results using multistage ensemble ANFIS FCM of 5 neural networks with the actual measured heating energy consumption data for the training and test period, respectively.

# 9 Conclusion

For the prediction of heating energy consumption in NTNU campus Gløshaugen, 50 different FFNNs are trained based on the coldest period in years 2009, 2010 and 2011 (318 samples), and tested for year 2012 (100 samples). Improvement in prediction accuracy using neural network ensemble is investigated. The main task in this method is achieving both accuracy and diversity of ensemble members. The accuracy is obtained by using adequate training algorithm and selecting number of neurons in hidden layer by trial and error method. K-means, as one of the most used clustering technique is used for separating trained networks into groups (clusters), and the best network in each cluster is selected for the ensemble member. Members are then aggregated into ensemble using various techniques: conventional methods (simple and weighted averaging) and multistage. Averaging the predictions of these networks resulted in an improvement in accuracy over the predictions of the best trained individual FFNN. Further improvement is obtained by training new neural network to combine the predictions of the original networks. In second level two different neural networks are analyzed: FFNN and ANFIS. Different ANFIS models are constructed using various identifi-

The accuracy of the proposed models in terms of coefficient of determination (R$^2$), while varying ensemble

cation methods: different membership functions (trimf, gbellmf, gaussmf), fuzzy C-means clustering (FCM) and subtractive clustering (SUB). All ensembles are trained and tested for various number of clusters. Multistage model, using ANFIS in second level is proven to be most effective. In this paper we have demonstrated that multistage ensembles, where the adaptive properties of a second layer network are used to combine the outputs of the individual ensemble members, offer enhanced performance over conventional combining methods and best trained single network.

## Acknowledgments

## References

AgroMetbase. *Weather database.* http://lmt.bioforsk.no/agrometbase/getweatherdata.php, 2014.

Bergesen, B., Groth, L. H., Langseth, B., Magnussen, I. H., Spilde, D., and Toutain, J. E. W. Energy consumption 2012-household energy consumption. Technical report, Technical Report 16, Norwegian Water Resources and Energy Directorate, 2013.

Bezdek, J. C. *Pattern recognition with fuzzy objective function algorithms.* Kluwer Academic Publishers, 1981.

Breiman, L. Bagging predictors. *Machine learning*, 1996. 24(2):123–140. doi:10.1007/BF00058655.

Chiu, S. L. Fuzzy model identification based on cluster estimation. *Journal of intelligent and Fuzzy systems*, 1994. 2(3):267–278.

Cibulka, J., Ebbesen, M. K., Hovland, G., Robbersmyr, K. G., and Hansen, M. R. A review on approaches for condition based maintenance in applications with induction machines located offshore. *Modeling, identification and control*, 2012. 15(3):191–203. doi:10.4173/mic.2012.2.4.

Council., E. Directive 2010/31/eu of the european parliament and of the council of 19 may 2010 on the energy performance of buildings. *Official Journal of the European Union 2010*, 2010. (18):13–35.

Dombaycı, Ö. A. The prediction of heating energy consumption in a model house by using artificial neural networks in denizli–turkey. *Advances in Engineering Software*, 2010. 41(2):141–147. doi:10.1016/j.advengsoft.2009.09.012.

Dunn, J. C. *A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.* Taylor & Francis, 1973.

Ekici, B. B. and Aksoy, U. T. Prediction of building energy consumption by using artificial neural networks. *Advances in Engineering Software*, 2009. 40(5):356–362. doi:10.1016/j.advengsoft.2008.05.003.

Ekonomou, L. Greek long-term energy consumption prediction using artificial neural networks. *Energy*, 2010. 35(2):512–517. doi:10.1016/j.energy.2009.10.018.

EnergyRemoteMonitoring. *ERM.* http://erm.tac.com/erm/, 2014.

Fu, Q., Hu, S.-X., and Zhao, S.-Y. A pso-based approach for neural network ensemble. *Journal of Zhejiang University (Engineering Science)*, 2004. 38(12):1596–1600.

Granitto, P. M., Verdes, P. F., and Ceccatto, H. A. Neural network ensembles: evaluation of aggregation algorithms. *Artificial Intelligence*, 2005. 163(2):139–162. doi:10.1016/j.artint.2004.09.006.

Hansen, L. K. and Salamon, P. Neural network ensembles. *IEEE transactions on pattern analysis and machine intelligence*, 1990. 12(10):993–1001. doi:10.1109/34.58871.

Ilić, S. A., Vukmirović, S. M., Erdeljan, A. M., and Kulić, F. J. Hybrid artificial neural network system for short-term load forecasting. *Thermal Science*, 2012. 16(suppl. 1):215–224. doi:10.2298/TSCI120130073I.

Jang, J.-S. Anfis: adaptive-network-based fuzzy inference system. *Systems, Man and Cybernetics, IEEE Transactions on*, 1993. 23(3):665–685. doi:10.1109/21.256541.

Johannessen, T. Varmeutviklingen i bygninger og klimaet. 1956.

Karatasou, S., Santamouris, M., and Geros, V. Modeling and predicting building's energy use with artificial neural networks: Methods and results. *Energy and Buildings*, 2006. 38(8):949–958. doi:10.1016/j.enbuild.2005.11.005.

Kumar, R., Aggarwal, R., and Sharma, J. Energy analysis of a building using artificial neural network: A review. *Energy and Buildings*, 2013. 65:352–358. doi:10.1016/j.enbuild.2013.06.007.

Kusiak, A., Li, M., and Zhang, Z. A data-driven approach for steam load prediction in buildings. *Applied Energy*, 2010. 87(3):925–933. doi:10.1016/j.apenergy.2009.09.004.

Lazarevic, A. and Obradovic, Z. Effective pruning of neural network classifier ensembles. In *Neural Networks, 2001. Proceedings. IJCNN'01. International Joint Conference on*, volume 2. IEEE, pages 796–801, 2001.

Li, K., Su, H., and Chu, J. Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: A comparative study. *Energy and Buildings*, 2011. 43(10):2893–2899. doi:10.1016/j.enbuild.2011.07.010.

MacQueen, J. et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1. Oakland, CA, USA., pages 281–297, 1967.

Melin, P., Soto, J., Castillo, O., and Soria, J. A new approach for time series prediction using ensembles of anfis models. *Expert Systems with Applications*, 2012. 39(3):3494–3506. doi:10.1016/j.eswa.2011.09.040.

Navone, H. D., Verdes, P. F., Granitto, P. M., and Ceccatto, H. A. Selecting diverse members of neural network ensembles. In *Neural Networks, 2000. Proceedings. Sixth Brazilian Symposium on*. IEEE, pages 255–260, 2000.

Opitz, D. W. and Shavlik, J. W. Actively searching for an effective neural network ensemble. *Connection Science*, 1996. 8(3-4):337–354. doi:10.1080/095400996116802.

Perera, D. W. U., Pfeiffer, C., and Skeie, N.-O. Modelling the heat dynamics of a residential building unit: Application to norwegian buildings. *Modeling, Identification and Control*, 2014. 35(1):43–57. doi:10.4173/mic.2014.1.4.

Qiang, F., Shang-Xu, H., and Sheng-Ying, Z. Clustering-based selective neural network ensemble. *Journal of Zhejiang University SCIENCE A*, 2005. 6(5):387–392.

Schapire, R. E. The strength of weak learnability. *Machine learning*, 1990. 5(2):197–227. doi:10.1007/BF00116037.

Sharkey, A. J. Multi-net systems. In *Combining artificial neural nets*, pages 1–30. Springer, 1999. doi:10.1007/978-1-4471-0793-4_1.

Singh, R., Kainthola, A., and Singh, T. Estimation of elastic constant of rocks using an anfis approach. *Applied Soft Computing*, 2012. 12(1):40–45. doi:10.1016/j.asoc.2011.09.010.

Siwek, K., Osowski, S., and Szupiluk, R. Ensemble neural network approach for accurate load forecasting in a power system. *International Journal of Applied Mathematics and Computer Science*, 2009. 19(2):303–315. doi:10.2478/v10006-009-0026-2.

Skullestad, A., Olsen, K., Rennehvammen, S., and Fløystad, H. Control of a gravity gradient stabilised satellite using fuzzy logic. *Modeling, Identification and Control*, 2001. 22(3):141–152. doi:10.4173/mic.2001.3.2.

Sretenovic, A. *Analysis of energy use at university campus*. M. sc. thesis, Norwegian University of Science and Technology, Department of Energy and Process Engineering, 2013.

Takagi, T. and Sugeno, M. Fuzzy identification of systems and its applications to modeling and control. *Systems, Man and Cybernetics, IEEE Transactions on*, 1985. (1):116–132. doi:10.1109/TSMC.1985.6313399.

Taylor, J. W. and Buizza, R. Neural network load forecasting with weather ensemble predictions. *Power Systems, IEEE Transactions on*, 2002. 17(3):626–632. doi:10.1109/MPER.2002.4312413.

Tveito, T. Heating degree-days - present consitions and scenario for the period 2021-2050. 2002.

Wojdyga, K. An influence of weather conditions on heat demand in district heating systems. *Energy and Buildings*, 2008. 40(11):2009–2014. doi:10.1016/j.enbuild.2008.05.008.

Zadeh, L. A. The role of fuzzy logic in modeling, identification and control. *Modeling, identification and control*, 1994. 15(3):191–203. doi:10.4173/mic.1994.3.9.

Zhang, G. P., Berardi, V., et al. Time series forecasting with neural network ensembles: an application for exchange rate prediction. *Journal of the Operational Research Society*, 2001. 52(6):652–664. doi:10.1057/palgrave.jors.2601133.

Zhou, Z.-H., Wu, J., and Tang, W. Ensembling neural networks: many could be better than

all. *Artificial intelligence*, 2002. 137(1):239–263. doi:10.1016/S0004-3702(02)00190-X.

Zhou, Z.-H., Wu, J.-x., Tang, W., and Chen, Z.-Q. Combining regression estimators: Ga-based selective neural network ensemble. *International Journal of Computational Intelligence and Applications*, 2001. 1(04):341–356. doi:10.1142/S1469026801000287.