Short-Term Transmission-Loss Forecast for the Slovenian Transmission Power System Based on a Fuzzy-Logic Decision Approach

Matej Rejc, Miloš Pantoš

Abstract— **In a deregulated environment, system operators are required to procure certain ancillary services, which, among others, may include compensation for active-power losses. This compensation usually involves long-term energy purchases and additional short-term energy purchases to cover the daily fluctuations. The short-term energy purchases require an accurate and quick short-term forecasting method that has to be efficiently applicable in day-ahead markets. This paper presents a novel short-term active-power-loss forecast method using power-flow analysis for the forecasted day. Specifically, this includes short-term load and generation forecasts as well as network-topology forecasts, which are used for the power-flow calculations and the resulting active-power loss calculations. To minimize the forecast errors, a fuzzy-weight grouping of the different short-term load and generation forecast results is proposed. An additional step for input-data pre-processing is presented, where the fuzzy clustering considers the patterns for training the forecasting models. The proposed approach was verified by using real data for the ENTSO-E interconnection and tested for the Slovenian power system. The forecasting results demonstrate the improved accuracy of the proposed approach.**

***Index Terms*—Clustering methods, Fuzzy logic, Short-term generation forecasting, Short-term load forecasting, Short-term loss forecasting.**

I. NOMENCLATURE

Indices:

ANN	Superscript index for artificial neural network method
F	Subscript index for forecasted day
MA	Superscript index for moving-average method
MLR	Superscript index for multiple-linear-regression method
d	Subscript index for day-type
e	Subscript index for load/generation curve in a subset
h	Subscript index for hour
i	Subscript index of error interval margin
j, *c*	Subscript index for cluster
n	Subscript index of most recent load/generation curve in a subset
r	Subscript index of cluster with highest degree of membership to the most recent load/generation curve in a subset
s	Subscript index for season

Variables and functions:

A	Membership function of AFE
AFE	Absolute forecast percentage error
B	Degree of membership
C	Hourly cloud cover
EMI	Error membership index
J	Fuzzy clustering objective function
SA	Surface of upper-right quadrant
SAF	Surface, representing the accuracy of a forecast method
SI	Surface of lower-left quadrant
SIF	Surface, representing the inaccuracy of a forecast method
T	Hourly temperature
a	Value of membership function of *A* at a specified error interval margin *r*
e	Random error
n	Most recent load/generation curve in a subset
p	Value of nodal injection of active power
r	Error interval margin
w	Fuzzy weight
x	Value of load/generation
α	Constant regression coefficient
β	Temperature regression coefficient
γ	Cloud-cover regression coefficient
δ	Similarity membership weight regression coefficient

Parameters and Constants:

NC	Number of clusters
NE	Number of load/generation curves in a subset
NI	Number of intervals
NH	Number of hours
NM	Number of load/generation curves considered in moving average calculation
NN	Number of load/generation nodal injections
NS	Number of seasons
m	Degree of fuzziness

M. Rejc is with the Laboratory of Power Systems, University of Ljubljana, Faculty of Electrical Engineering, SI-1000 Ljubljana, Slovenia (e-mail: matej.rejc@fe.uni-lj.si).

M. Pantoš is with the Laboratory of Power Systems, University of Ljubljana, Faculty of Electrical Engineering, SI-1000 Ljubljana, Slovenia (e-mail: milos.pantos@fe.uni-lj.si).

Matrices and Vectors:

B	Degree of membership matrix
C	Hourly cloud-cover matrix
EMI	Error membership index matrix
T	Hourly temperature matrix
a	Membership value matrix
e	Matrix of random errors
p	Matrix of nodal injections of active power
r	Error interval margin matrix
v	Cluster center matrix
w	Matrix of fuzzy weights
x	Historical load/generation curve matrix

Abbreviations:

AC	Alternating current
AFE	Absolute percentage error
ANN	Artificial neural network
EMI	Error margin index
MA	Moving average
MLR	Multiple linear regression
SMW	Similarity-membership weight

II. Introduction

THE restructuring of the electric-power industry introduced competition to the generation of power, the concept of electricity supply as a commodity market, and new roles for the utilities. One such aspect defined by new regulations is the procurement and compensation of ancillary services within a competitive market. Ancillary services in the supply of electricity are defined as services essential for the functioning of electric power systems and, among other services, may include compensation for active-power losses [1], [2]. These losses must be compensated by the system operator. This requires the system operator to purchase the required energy, which is usually done in several stages, most often as yearly, monthly and daily volume purchases [2]. In the event the purchased energy does not cover the actual active-power losses, secondary active-power reserve must be activated. This is not cost effective and may in a worst-case scenario threaten the security of the system due to the fact that the secondary active-power reserve is primarily intended to maintain or establish a balance between generation and consumption in the control area in line with the forecasted operation schedules and it restores the regulative scope of primary control. For this reason, the cost of secondary active-power reserve energy is usually higher than the cost of energy required to compensate active-power losses bought on a day-ahead market, due to the fact that secondary active-power reserve includes energy purchases and control-output time-range specifications. The increasing economic pressure on system operators thus requires the use of accurate forecasting tools to forecast the active-power losses in order to be efficiently applicable in day-ahead markets.

Short-term active-power-loss forecasting is closely related to short-term forecasts of the system loading conditions, the imports and exports of electrical energy, i.e., the power transits, and the network topology. These system conditions define the power flows and the losses throughout the system.

The numerous short-term forecasting tools applied in the field of power engineering, such as load forecasting, can be applied for direct active-power-loss forecasting. However, due to the nonlinear behavior of active-power losses and the numerous impact factors, mathematical forecast models are extremely hard to define directly. Taking this into consideration, the idea is to forecast the losses indirectly using a power-flow analysis of the forecasted system states, i.e., the network topology, the system load, the generation and the power transit. Network-topology forecasts are defined according to the system operator’s scheduled maintenance plans, while the system load, generation and power flow transits are forecasted using short-term forecasting tools.

These short-term forecasting tools can be divided into two groups: the statistical methods and the artificial-intelligence-based forecasting methods [3]. The first group includes multiple linear-regression methods [4], [5], time series methods, such as the autoregressive and moving average methods [6], [7], the general exponential smoothing method [8], the method based on state-space models [9], and the support-vector-regression method [10], [11]. The statistical methods predict the response variable of ordinary days very well, but they can give unsatisfactory results for holidays and other anomalous days, due to their inflexibility [12].

The second group of forecasting methods incorporates methods that can be interpreted as adaptive machines learning algorithms. The members of this group are methods based on artificial neural networks (ANNs) [13], [14], expert system methods [15], and fuzzy inference [16]. The artificial-intelligence based methods are capable of finding the nonlinear relationship between the response variables and its impact factors. However, they have shortcomings, which can lie in over-fitting and long training times for ANNs and the difficulty of expressing intuitive expert knowledge and fuzzy rules for expert system methods and fuzzy inference.

Considering the nonlinear nature of active-power losses and its impact factors, this paper presents a novel approach to short-term active-power-loss forecasting based on the power-flow calculations of forecasted system states using short-term load and generation forecasts and network-topology forecasts. The proposed approach integrates three different forecast methods: the weather-sensitive ANN approach, the multiple linear regression method and the non-weather-sensitive moving-average method. The reason behind multiple forecasting methods is the complexity of the problem and the requirement of a large amount of information and data, which is often partly missing. In addition, a particular method could give unsatisfactory results, e.g. due to missing data, so the main question addressed here is how to make a robust and reliable method that obtains loss forecasts with acceptable accuracy.

The solution, i.e., the load and generation forecast, is obtained using the proposed fuzzy-weight decision approach in order to minimize the forecast errors of unsatisfactory individual forecasting methods and thus increase the robustness of the forecast procedure. The forecasts are then followed by AC power-flow calculations for the forecasted day, which capture the physics of the power flows and the resulting active-power losses. This may reduce the forecast error compared to direct transmission-loss forecasting methods and the system operator can more easily discern the effect of various loss impact factors. The proposed procedure does not present an additional burden on the system operator, as the needed data is already available to the system operator and power-flow analysis must be performed to ensure the safe operation of the power system on a daily basis. The presented approach represents an example of a practical implementation that is currently being designed for the use in daily operations by the Slovenian system operator.

This paper is organized as follows: Section III presents the transmission-loss forecast methodology and is further divided into subsections III.A–III.F. Subsection III.A presents the input-data analysis and the fuzzy clustering of the training patterns. Subsection III.B presents the short-term load forecasting procedure, i.e., the multiple linear regression method, the moving-average and the ANN approach and the fuzzy integration of the forecasting results in detail. The generation and network-topology forecasts are described in subsections III.C and III.D, respectively. The interconnected power systems' influence on the forecasts are described in subsection III.E and the transmission-loss forecast is described in subsection III.F. Section IV presents the results of the presented methodology for the Slovenian power system using the model of the ENTSO-E power system. Finally, the conclusions are given in Section V.

III. Transmission-Loss Forecast Methodology

Active-power losses are mainly dependent on impact factors that can be classified according to their influence on:

- the system load, i.e., the weather, social and time/seasonal factors,
- the system generation that balances the system load based on an appropriate unit commitment and an economic dispatch,
- the network-topology,
- the influence of the surrounding power systems' imports/exports, i.e., electricity costs, lack of generation capabilities, etc..

Although system loading conditions can have a high correlation with transmission losses in weakly interconnected power systems, using load forecast results only to directly forecast active-power losses in highly interconnected transmission power systems can decrease the accuracy of the forecast procedure due to cross-border power transits and additionally, network specifics topology can greatly affect transmission losses. The proposed transmission-loss forecast methodology, Fig. 1, is based on an assessment of the influences of the described impact factors on the active-power losses that apply:

- the preparation of the input data for forecasting model fitting,
- the short-term load forecast,
- the short-term generation forecast,
- the network-topology forecast,
- the active-power loss forecast.

The first task is the input-data analysis that in its first step performs a similar-day grouping procedure within seasons in order to capture the differences in the daily and seasonal load profiles, [11]. After these data subsets are defined, the next step is data clustering within the subsets that differentiates the subset data into clusters corresponding to the calculated fuzzy weights. For this purpose the data fuzzy c-means clustering method [17] is applied. This is one of the most commonly used methods for fuzzy clustering and has already been used in the field of short-term load forecasting [18], [19]. The complete procedure is presented in detail in subsection III.A.

The second task is the short-term load and generation forecasts that are performed separately, applying three different forecast methods:

- the multiple linear regression method,
- the moving average method,
- the ANN method.

The reason behind multiple forecasting methods is the complexity of the problem and the requirement of a large amount of information and data, which is often partly missing. Also a particular method could give unsatisfactory results, e.g. due to missing data, thus three methods are combined in order to improve the forecast accuracies and robustness.

The proposed approach combines the forecasts using fuzzy logic based on their respective fuzzy weights. The fuzzy weights are determined based on the methods' previous successes. The procedure for load forecasting is presented in detail in subsection III.B and the proposed fuzzy-weight-based grouping in subsection III.B.4. The generation forecasting is presented in subsection III.C.

The third task is the network-topology forecasts, which is presented in subsection III.D, followed by the final task, i.e., the transmission-loss forecast based on AC power-flow calculations of the power-flow model with the forecasted system states. The transmission-loss forecast is presented in subsection III.F.

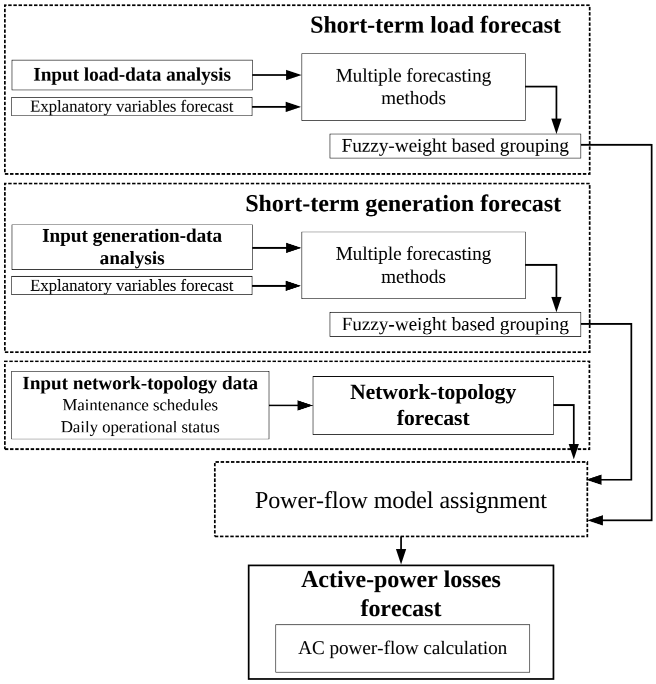


Fig. 1. Flow chart for the transmission-losses forecast method.

A. Input-data analysis

Methods applied in the load and generation forecast are based on historical data, i.e., the daily system load/generation curves, over a longer time period. Due to large deviations of the load/generation within that period a similar-day grouping procedure within seasons and day-types is required in order to differentiate the daily and seasonal load/generation profiles. If yearly data or data for a longer period is available, it is reasonable to define four data subsets for each season and five data subsets for different day-types: Mondays, Tuesdays-Thursdays, Fridays, Saturdays and Sundays with holidays [11] resulting in a total of 20 data subsets. This approach limits the historical data to the most relevant and representative data for the forecast model fitting.

Further, for each subset the fuzzy-weight clustering of daily load and/or generation curves is performed in order to differentiate the data into clusters corresponding to the calculated fuzzy weights. In this way, each load and/or generation daily curve impacts the forecasted load and/or the generation curve according to its fuzzy weight. This approach gives more importance to the most recent data, which receives higher fuzzy weights.

For this purpose the fuzzy c-means clustering method [17] is applied. Fuzzy clustering is used to find fuzzy clusters and the associated cluster centers by which the structure of the data within the subset is represented as the best possible. This is based on a minimization of the objective function in season s and for a day type d, $J_{s,d}$:

$$J_{s,d} = \sum_{e=1}^{NE} \sum_{c=1}^{NC} B_{s,d,e,c}^{m} \cdot \left\| \mathbf{x}_{s,d,e,(.)} - \mathbf{v}_{s,d,c,(.)} \right\|^2 , \qquad (1)$$

where $\mathbf{x}_{s,d,e,(.)} = [x_{s,d,e,1}, x_{s,d,e,2}, \ldots, x_{s,d,e,h}, \ldots, x_{s,d,e,NH}]^{\mathrm{T}}$ is the e-th daily system load/generation curve in season s and for the day type d. NE, NC, NH and m represent the number of load/generation curves in the subset, the number of clusters, the amount of hourly data in a daily curve, and the degree of fuzziness, respectively. The degree of membership in season s, for day type d, of load/generation daily curve e in the cluster c, $B_{s,d,e,c}$, takes the value between 0 and 1, where the constraints represent no membership and full membership to the cluster c. It is calculated as [17]:

$$B_{s,d,e,c} = \left(\sum_{j=1}^{NC} \left(\frac{\left\| \mathbf{x}_{s,d,e,(.)} - \mathbf{v}_{s,d,c,(.)} \right\|^2}{\left\| \mathbf{x}_{s,d,e,(.)} - \mathbf{v}_{s,d,j,(.)} \right\|^2} \right)^{\frac{1}{m-1}} \right)^{-1} , \qquad (2)$$

where the cluster center in season s, for day type d and cluster c, $\mathbf{v}_{s,d,c,(.)} = [v_{s,d,c,1}, v_{s,d,c,2}, \ldots, v_{s,d,c,h}, \ldots, v_{s,d,NC,NH}]^{\mathrm{T}}$, is calculated as:

$$\mathbf{v}_{s,d,c,(.)} = \frac{\sum_{e=1}^{NE} B_{s,d,e,c}^{m} \cdot \mathbf{x}_{s,d,e,(.)}}{\sum_{e=1}^{NE} B_{s,d,e,c}^{m}} . \qquad (3)$$

The fuzzy c-means algorithm is carried out through an iterative optimization with the update of the degree membership $B_{s,d,e,c}$ and the cluster center $\mathbf{v}_{s,d,c,(.)}$. The value of the fuzziness index m was chosen as 2, as it suits the majority of problems and reduces the computation time [17]. The Euclidian distance norm was used to express the distances between the load curves and the cluster centers.

1) Derivation of the similarity-membership weights

After the fuzzy clustering is performed the degree of memberships of all the load/generation curves in the analyzed data subset for all the clusters are obtained. Furthermore, it is assumed that the most recent load/generation curve in the subset is the most relevant for forecasting. From the set of the degree of memberships of the last load/generation curve n, $\mathbf{B}_{s,d,n,(.)} = [B_{s,d,n,1}, B_{s,d,n,2}, \ldots, B_{s,d,n,r}, \ldots, B_{s,d,n,NC}]^{\mathrm{T}}$, it is easy to identify the cluster r, in which the concerned load/generation curve n has the highest degree of membership value, denoted as $B_{s,d,n,r}$. The degree of memberships of all the load/generation curves in that particular cluster r, $\mathbf{B}_{s,d,(.),r} = [B_{s,d,1,r}, B_{s,d,2,r}, \ldots, B_{s,d,NE,r}]^{\mathrm{T}}$, represents the similarity membership weights (SMWs), Fig. 2. The SMWs are used in the forecasting procedures presented in subsections III.B and III.C. In this way, more importance is given to the data most similar to the recent data n in the forecasting procedure, since the forecasted day is expected to have higher degrees of membership compared to the cluster r. The non-similar data are not discarded entirely as they are also included in the forecasting procedure with the appropriate degree of memberships that are expected to have lower values.

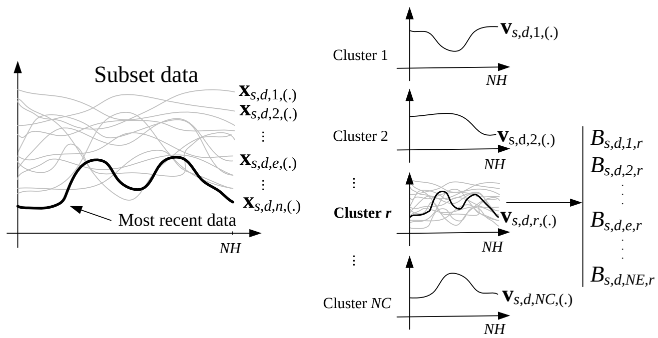


Fig. 2. Derivation of the similarity membership weights.

B. *Short-term load forecasting*

In the proposed short-term load-forecasting method, three different methods, i.e., the multiple linear regression method, the moving average method, and the ANN method, are combined using fuzzy-weight grouping, as explained in the subsequent text.

1) *Multiple Linear Regression*

The multiple linear regression method is an effective method for formulating the linear relationship between the response variable and the explanatory variables, i.e., the impact factors. In this paper, the following explanatory variables were used:

- the hourly temperature,
- the hourly cloud cover,
- the SMWs explained in III.A.1.

The model is represented as a linear function of the explanatory variables and regression coefficients:

$$\begin{aligned}\mathbf{x}_{s,d,(.),h} &= \alpha_{s,d,h} + \beta_{s,d,h}\mathbf{T}_{s,d,(.),h} + \gamma_{s,d,h}\mathbf{C}_{s,d,(.),h} + \\ &+ \delta_{s,d,h}\mathbf{B}_{s,d,r,(.)} + \mathbf{e}_{s,d,(.),h}\end{aligned}, \quad (4)$$

where $\mathbf{x}_{s,d,(.),h} = [x_{s,d,1,h},\ x_{s,d,2,h},\ \ldots,\ x_{s,d,e,h},\ \ldots,\ x_{s,d,NE,h}]^{\mathrm{T}}$ is the vector of all the *NE* system loads in season *s*, for day type *d*, for hour *h*. The coefficients $\alpha_{s,d,h}$, $\beta_{s,d,h}$, $\gamma_{s,d,h}$, and $\delta_{s,d,h}$ represent the constant regression coefficient, the temperature regression coefficient, the cloud-cover regression coefficient, and the SMW regression coefficient in season *s*, for day type *d*, for hour *h*, respectively. $\mathbf{T}_{s,d,(.),h} = [T_{s,d,1,h},\ T_{s,d,2,h},\ \ldots,\ T_{s,d,e,h},\ \ldots,\ T_{s,d,NE,h}]^{\mathrm{T}}$ is the vector of all the *NE* temperatures in season *s*, for day type *d*, for hour *h* and $\mathbf{C}_{s,d,(.),h} = [C_{s,d,1,h},\ C_{s,d,2,h},\ \ldots,\ C_{s,d,e,h},\ \ldots,\ C_{s,d,NE,h}]^{\mathrm{T}}$ is the vector of all the *NE* cloud cover data in season *s*, for day type *d*, for hour *h*. The vector $\mathbf{e}_{s,d,(.),h} = [e_{s,d,1,h},\ e_{s,d,2,h},\ \ldots,\ e_{s,d,e,h},\ \ldots,\ e_{s,d,NE,h}]^{\mathrm{T}}$ consists of all the *NE* random errors in season *s*, for day type *d*, for hour *h*.

The least-squares method is applied to minimize the random-error term and to determine the regression coefficients from the historical data. Once these parameters are obtained, the load at hour *h* of the forecasted day, i.e., day *NE*+1, in season *s*, for day type *d*, $x_{s,d,NE+1,h}$, can be forecasted as:

$$\begin{aligned}x_{s,d,NE+1,h} &= \alpha_{s,d,h} + \beta_{s,d,h}T_{s,d,NE+1,h} + \\ &+ \gamma_{s,d,h}C_{s,d,NE+1,h} + \delta_{s,d,h}\cdot\max\left(\mathbf{B}_{s,d,(.),r}\right)\end{aligned}, \quad (5)$$

where the maximum value of SMW in $\mathbf{B}_{s,d,(.),r}$ is used for the forecasted day, i.e., *NE*+1, as it is presumed that the system-load daily curve will take the shape of the system load curve with the highest degree of membership to the cluster *r*, in season *s*, for day type *d*.

2) *Moving average model*

The most fundamental time series model is the moving average model, and represents a non-weather sensitive method. The idea behind using a non-weather sensitive method is to increase the robustness of the forecast procedure in cases where the historical weather data or weather forecasts are missing. The moving average model is defined as:

$$\mathbf{x}_{s,d,NE+1,(.)} = \frac{\sum_{e=NE-NM}^{NE} B_{s,d,e,r}\cdot\mathbf{x}_{s,d,e,(.)}}{\sum_{e=NE-NM}^{NE} B_{s,d,e,r}}, \quad (6)$$

where $\mathbf{x}_{s,d,NE+1,(.)} = [x_{s,d,NE+1,1},\ x_{s,d,NE+1,2},\ \ldots,\ x_{s,d,NE+1,h},\ \ldots,\ x_{s,d,NE+1,NH}]^{\mathrm{T}}$ is the vector of hourly values of the system load in the forecasted day *NE*+1, in season *s*, for day type *d*. Since the method is based on a moving average, *NM* represents the number of system daily-load curves considered in the calculation.

3) *Artificial Neural Networks*

Due to the ability to model the nonlinear relationships between the load and the various impact factors, ANNs are one of the most popular techniques used for short-term load forecasting. For the purpose of this paper, different ANN structures were tested to optimize the effectiveness of the used ANN model. A two-layer feed-forward network with a back-propagation algorithm was used to define the functional relationship between the load and the following impact factors:

- the hourly temperature,
- the cloud cover,
- the SMWs, explained in III.A.1.

The neural model used had one hidden layer with sigmoid neurons and one output layer with linear output neuron. Detailed formulations and descriptions of the ANN structures can be found in [13], [14].

4) *Fuzzy-weight-based grouping*

The applied forecast methods have some unique advantages and disadvantages that have to be considered, e.g., the inflexibilities of statistical methods and the lack of ability to forecast the load for holidays and other anomalous days, the problem of over-fitting and under-fitting of the ANN [12], [14], the effect of incomplete or inaccurate weather forecasts on weather-sensitive methods, etc. In order to improve the forecast precision and robustness, the results of the applied forecast methods are combined. A simple way is to average the results, which can lower the risk of an unsatisfactory forecast for an individual method, or it can negatively affect the precision in the case that one or more methods repeatedly give less accurate results than others.

In this paper, a fuzzy weight-based method is proposed to combine the forecast results of the forecasting methods used. This requires a historical analysis of the applied forecast methods and their absolute forecast percentage errors (AFEs):

$$AFE_{s,d,e,h} = \frac{\sqrt{\left(x_{F,s,d,e,h} - x_{s,d,e,h}\right)^2}}{x_{s,d,e,h}}, \qquad (7)$$

where $x_{F,s,d,e,h}$ is the forecasted system load in season s, for day type d, for daily load curve e, for hour h that is compared with the actual system load $x_{s,d,e,h}$.

Furthermore, a fuzzy weight, i.e., the error membership index (*EMI*) for season s, for day type d, for daily load curve e, for hour h, $EMI_{s,d,e,h}$, is calculated as:

$$EMI_{s,d,e,h} = A_{s,d,e,h}(AFE_{s,d,e,h}) = \\ = \begin{cases} 1 & AFE_{s,d,e,h} \le r_{s,d,e,h,1} \\ K_{s,d,e,h,i} & r_{s,d,e,h,i} < AFE_{s,d,e,h} \le r_{s,d,e,h,i+1} \\ 0 & r_{s,d,e,h,NI} < AFE_{s,d,e,h} \end{cases}, \qquad (8)$$

$$K_{s,d,e,h,i} = \frac{a_{s,d,e,h,i} - a_{s,d,e,h,i-1}}{r_{s,d,e,h,i} - r_{s,d,e,h,i-1}}\left(AFE_{s,d,e,h} + 1\right) + a_{s,d,e,h,i},$$

where $a_{s,d,e,h,i}$ is the membership value at the margin i of the membership function of *AFE* $A_{s,d,e,h}$ at the error interval margin i, $r_{s,d,e,h,i}$, for season s, for day type d, for daily load curve e, for hour h. It is a piecewise linear function with empirically defined interval margins on the x and y axes. The membership function $A_{s,d,e,h}$. is presented in Fig. 3, where $\mathbf{a}_{s,d,e,h,(.)} = [a_{s,d,e,h,1}, a_{s,d,e,h,2}, \ldots, a_{s,d,e,h,NI}]^{\mathrm{T}}$, where $a_{s,d,e,h,1} = 0$ and $a_{s,d,e,h,\mathrm{NI}} = 1$, and $\mathbf{r}_{s,d,e,h,(.)} = [r_{s,d,e,h,1}, r_{s,d,e,h,2}, \ldots, r_{s,d,e,h,NI}]^{\mathrm{T}}$, where $r_{s,d,e,h,1} = 0$ and $r_{s,d,e,h,NI} = \infty$. The reason for the various intervals is to assign different fuzzy weights, $EMI_{s,d,e,h}$, to different error intervals $[r_{s,d,e,h,i}, r_{s,d,e,h,i-1}]$.

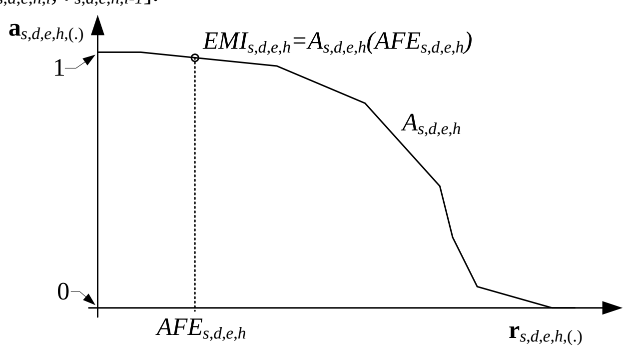
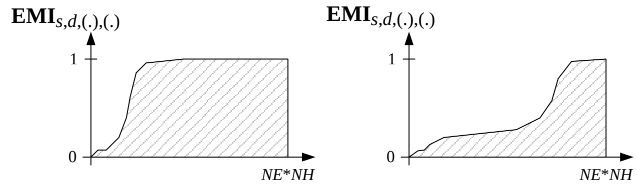


Fig.3: Membership function $A_{s,d,e,h}$.

The membership function $A_{s,d,e,h}$ has small gradients on the intervals where *EMI* is close to 1 or 0 as the difference between an extremely high and very high or an extremely low and very low accuracy is not as important. In both cases *EMI* is close to 1 or 0, respectively. The membership function has higher gradients at values in between. This enables a higher drop-off rate of the *EMI* values as *AFE* increases in a certain interval, e.g., the difference between an *AFE* of 3 % and 8 % is very important, so the *EMI* value drops at a higher rate, as where the difference between 40 % and 51 % *AFE* is not. The method with such a high AFE is highly inaccurate in both the latter cases and consequently the *EMI* is close to 0.

Once the *EMI*s are calculated for all the past *NE* forecasts, and all hours, *NH*, they are sorted in ascending order for every observed hour h. If the assessed forecast method is very accurate, the corresponding forecast errors are very small, resulting in *EMI*s near to 1, Fig. 4a. In contrast, inaccurate forecast methods have lower *EMI*s, as presented in Fig. 4b.



a) Successful forecast method b) Unsuccessful forecast method

Fig 4. Sorted *EMI*s.

The values of the normalized surfaces of the sorted *EMI*s, Fig. 4, can be used to assign fuzzy weights to the individual methods. However, observations of the resulting fuzzy weights showed that methods with a relatively large number of mid or low accuracies and some high accuracies were given unreasonably high value weights. The observations showed that by observing the normalized surfaces only, the accuracies or inaccuracies of a method are not strictly rewarded or penalized.

For this reason, the following approach is proposed in this paper. The surface is further divided into four quadrants, where the surfaces $SA_{s,d}$ and $SI_{s,d}$ represent the upper-right and lower-left quadrants, respectively. The surfaces $SAF_{s,d}$ and $SIF_{s,d}$ represent how accurate and inaccurate the forecast method was in season s, for day type d, respectively, Fig. 5.

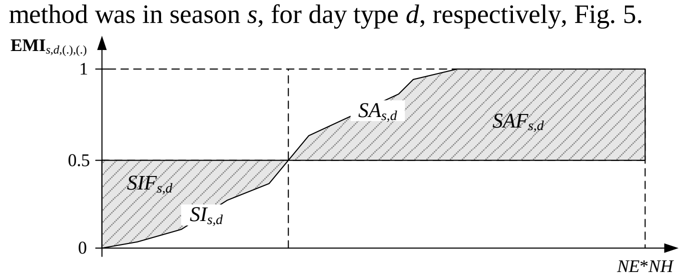


Fig 5. Division of sorted EMIs into two subgroups.

According to the two subgroups, $SAF_{s,d}$ and $SIF_{s,d}$, the fuzzy weight in season s on day type d, $w_{s,d}$, is calculated as:

$$w_{s,d} = \frac{\frac{1}{2}\left(1 + sign\left(SAF_{s,d} - SIF_{s,d}\right)\right)}{\max\left(SA_{s,d}, SI_{s,d}\right)}\left(SAF_{s,d} - SIF_{s,d}\right). \qquad (9)$$

Individual forecast methods are assigned fuzzy weights that belong to a fuzzy set to a greater or lesser degree represented by real-number values ranging in the interval [0, 1], respectively. The margin values 0 and 1 denote the absolutely inaccurate and absolutely accurate method.

The forecast results of the methods used are combined using the fuzzy weights:

$$\mathbf{w} = \begin{bmatrix} w_{s,d}^{MLR} \\ w_{s,d}^{MA} \\ w_{s,d}^{ANN} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_{s,d,NE+1,h}^{MLR} \\ x_{s,d,NE+1,h}^{MA} \\ x_{s,d,NE+1,h}^{ANN} \end{bmatrix}, \qquad (10)$$

$$x_{s,d,NE+1,h} = \frac{1}{w_{s,d}^{MLR} + w_{s,d}^{MA} + w_{s,d}^{ANN}} \mathbf{w}^{\mathrm{T}} \cdot \mathbf{x} \, . \tag{11}$$

where the superscripts *MLR*, *MA*, and *ANN* are assigned to the forecast methods applied in the methodology, i.e., the multiple linear regression method, the moving average method, and the ANN method. The results of the forecasts are assessed and the fuzzy weights are updated continuously.

Fig. 6 presents the complete procedure for the presented short-term load-forecasting methodology, composed of the input-data analysis, the short-term load forecasts using MLR, MA and the ANN method, fuzzy weight-based assignment and the resulting integration of the forecasting results.

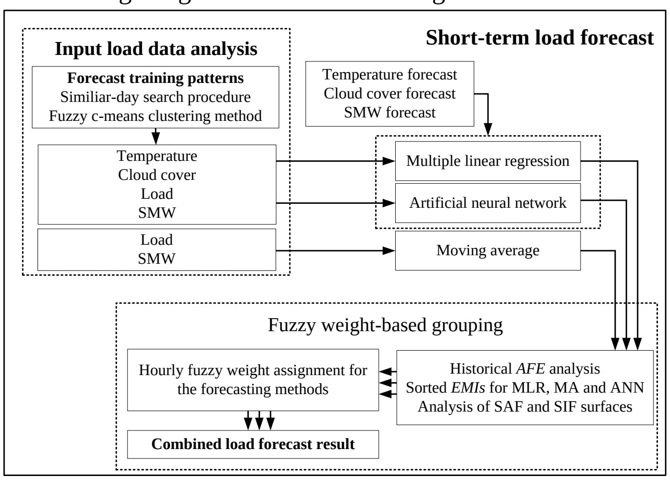


Fig. 6. Flow chart for the short-term load forecasting.

C. Generation forecasting

Generation forecasts and the resulting fuzzy weights are performed in the same way as load forecasts. Due to the limited data for electricity-market transactions between generators and its counterparties, only the following impact factors were considered in multiple linear regression and ANN:

- the precipitation for areas with large hydro-power integration,
- the wind speeds for areas with large wind-power integration,
- the load,
- the SMWs explained in III.A.1.

The available data used for the purpose of the paper was in the form of nodal injections for the observed interconnected power systems. Considering this, the areas with high wind penetration were forecasted individually with the associated wind speed data. Likewise, for areas with large hydro-power generation, precipitation was used as an added explanatory variable. In areas high thermal-power integration, precipitation and wind speeds were not used as explanatory variables.

The moving average forecast only used historical generation curves to forecast the generation. It should be noted that the short-term electricity prices are derivable from the considered impact factors. However, the electricity prices can be included as an additional explanatory variable, as well.

Fig. 7 presents the complete procedure for the presented short-term generation forecasting methodology, composed of the input-data analysis, the short-term generation forecasts using MLR, MA and the ANN method, fuzzy weight-based assignment and the resulting integration of the forecasting results.

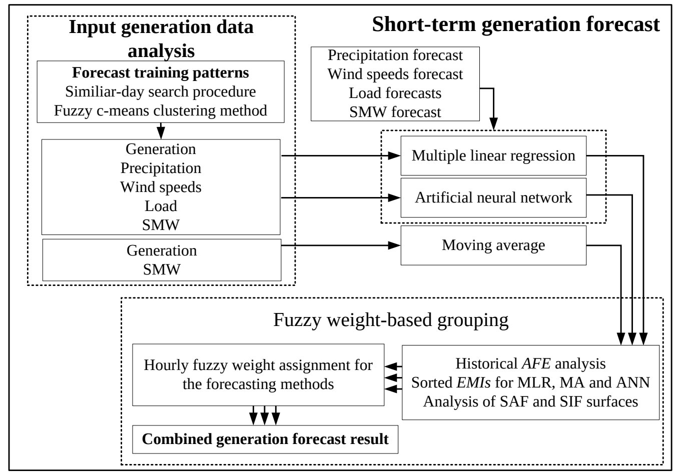


Fig. 7. Flow chart for the short-term generation forecasting.

D. Network-topology forecasting

The network-topology represents an important impact factor and must also be taken into account when forecasting losses. As active power losses are proportional to the square of the load current, any re-routing of power can drastically affect the losses.

The network-topology forecasting is performed using information based on the transmission system operators' maintenance schedules and the day-to-day system-operation status and are considered to be always accurate. Specifically, this takes into account yearly maintenance plans of power line switch-offs and refits and day-to-day system topology, i.e., switched off lines due to unexpected occurrences. The disconnected or reconnected power lines and transformers, i.e. the topology for the forecasted day is assigned to the power flow model with the logical switches setting the branch status.

E. Influence of interconnected power systems

With the introduction of transmission open access the amount of cross-border transactions drastically increased, power-flow transits additionally increase the power flows and indirectly increase the active-power losses. The presented methodology includes power-flow transits in the form of system load and the generation forecasts of foreign interconnected systems, as described in subsection III.B and III.C.

When the network-topology, load and generation forecasts for the interconnection are known, i.e., the procedures described in subsections III.A–III.E are concluded, the transmission losses can be calculated using a power-flow analysis, described in subsection III.F.

F. Transmission-loss calculation

The power-flow model must be composed of the analyzed interconnected transmission systems. The model is updated according to the load, generation and network configuration forecasts.

The forecasted aggregated load and generation are proportionally shifted compared to the load pattern of the most recent previous similar day, i.e., all nodal load injections in the forecasted area are proportionally shifted up or down in accordance with the ratio of previous similar-day system load and the forecasted system load for each hour. Load forecasts are described in subsection III.B. Similarly, the generation nodal injections are shifted proportionally in accordance with the ratio of previous similar-day system generation and the forecasted system generation for each hour. Generation forecasts are described in subsection III.C. Power flow model nodal injections are forecasted in accordance with:

$$\mathbf{p}_{s,d,NE+1,h,(.)} = \frac{x_{s,d,NE+1,h}}{x_{s,d,NE,h}} \cdot \mathbf{p}_{s,d,NE+1,h,(.)}, \tag{12}$$

where $\mathbf{p}_{s,d,NE+1,h,(.)} = [p_{s,d,NE+1,h,1},\ p_{s,d,NE+1,h,2},\ \ldots,\ p_{s,d,NE+1,h,NN}]^{\mathrm{T}}$ is the vector of all the NN forecasted nodal injections for season s, day type d, hour h, $\mathbf{p}_{s,d,NE,h,(.)} = [p_{s,d,NE,h,1},\ p_{s,d,NE,h,2},\ \ldots,\ p_{s,d,NE,h,NN}]^{\mathrm{T}}$ is the vector of all the NN known nodal injections for the previous similar-day NE for season s, day type d, hour h. $x_{s,d,NE+1,h}$ is the forecasted load/generation for season s, day type d, hour h and $x_{s,d,NE,h}$ is the load/generation for season s, day type d, hour h for the previous similar-day NE.

It is assumed that the nodal load/generation pattern does not vary greatly between similar days. In cases where specific load or generation patterns are known, such as in the case of long-term contracts for specific areas or generation units, the injections are changed accordingly, i.e., forecasts are performed for specific nodes and not for larger areas. The network-topology is modified in accordance with the network-topology forecast, i.e. power flow model branch logical switches are set accordingly. The network-topology forecast is described in subsection III.D.

The transmission losses are thus forecasted by applying AC power-flow calculations.

IV. Simulation Results

In order to test the proposed method, the UCTE region of the European synchronous transmission grid operated by ENTSO-E was investigated. The target area for the validation of the proposed transmission-loss forecast method was the Slovenian transmission system. The active-power loss-forecast procedure was tested and validated for a one-month period with the available data: hourly load and generation, hourly temperature, cloud cover, precipitation and wind speeds for a complete year. Individual method explanatory variables have been selected based on various tests and statistical relationships between the variables [12]. In addition, a correlation analysis of the various impact factors on transmission losses for the Slovenian power system was performed, with correlations between transmission losses and system loading and amount of power transits in regards to high or low power transits throughout the system, i.e. when power transits are below or above 50 % of entire system load. Table I shows the correlation analysis results, which confirm the validity of using an indirect forecasting procedure to improve forecast accuracy.

TABLE I
Correlation analysis between transmission losses and system loading and amount of power transits

	Transmission losses	
	Low transits	High transits
System load	0.63	0.59
Power transits	0.40	0.45

For the purpose of the paper, a one-year time-series of historical data was used to calibrate the forecast model. This was needed to take the seasonal changes into account and to ensure that enough data samples were used to fit the forecasting model.

Table II presents the membership values **a** of the membership function A at error interval margins **r**. The membership function A defined by **a** and **r** defines the *EMI* values for specific *AFE* values, Fig. 3.

TABLE II
Membership function A value and error intervals

$\boldsymbol{r}_{s,d,e,h,(.)}$	[0, 1]	[1, 3]	[3, 4]	[4, 5]	[5, 10]	[10, ∞]
$\boldsymbol{a}_{s,d,e,h,(.)}$	[1, 1]	[1, 0.8]	[0.8, 0.5]	[0.5, 0.2]	[0.2, 0]	[0, 0]

In order to assess the robustness of the proposed forecasting procedure, two case studies were performed, i.e., case-study A with complete data and case-study B with incomplete data. In the first case, complete weather, load and generation data are available, while in the second case, the weather data is partly unavailable.

For the purpose of training the ANN, a two-layer feed-forward network with a back-propagation algorithm with one hidden layer with 50 sigmoid neurons and one output layer with linear output neuron was applied. The selected ANN was trained with 70% of available data, and validated and tested using 15 % of the data each. The moving average model used 12 historical load/generation daily curves. The multiple linear regression used the same data as the ANN, i.e. all the historical data in the observed season and day-type cluster.

A. *Case study with complete data*

Table III shows the mean absolute percentage error (MAPE) and the standard deviation for the individual short-term load and generation forecast methods and the combined results for the Slovenian power system for the analyzed month. At certain time intervals, an individual method may give better results; however, it may also deviate widely from the actual load or generation at different time intervals. It was shown that the combined forecast minimizes such individual large deviation errors, while still retaining a high forecast precision of 2.02 MAPE and 0.82 standard deviation of MAPE for load forecasts, compared to the use of individual methods, where the MAPE ranged from 2.19 to 3.35 and the standard deviations of MAPE from 0.98 to 1.14 for load forecasts. For generation forecasts the combined forecast reached 8.05 MAPE and 11.26 standard deviation of MAPE while the individual method MAPE ranged from 9.12 to 12.9 and the standard deviation from 13.15 to 18.24.

TABLE III
MAPE AND STANDARD DEVIATION COMPARISON OF THE LOAD AND GENERATION FORECAST RESULTS FOR THE ANALYZED MONTH FOR THE SLOVENIAN POWER SYSTEM

	Load		Generation	
	MAPE	St. deviation (%)	MAPE	St. deviation (%)
Multiple linear regression	2.83	1.04	12.9	18.24
Moving average	3.35	1.14	9.12	13.15
ANN	2.19	0.98	9.34	13.55
Combined forecast	2.02	0.82	8.05	11.26

Fig. 8 shows an example of the day-ahead system-load forecasts for the Slovenian power system for a randomly chosen weekday and the MAPE of the forecasts for the day. The combined forecast has the lowest MAPE for the chosen day, with a MAPE of 0.9, compared to the MAPE of 2.52, 2.31 and 1.27 for the multiple linear regression, the moving average and the ANN forecast, respectively.

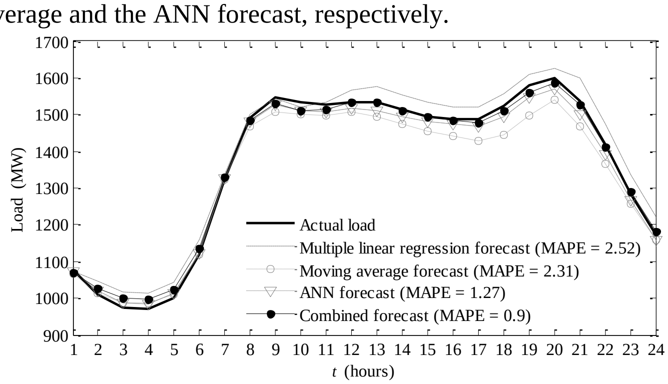


Fig. 8. Comparison of load forecasting results for the chosen day for the Slovenian power system.

Fig. 9 shows an example of previous membership values *EMI* for similar previous day-types for hour 16. In accordance with the described procedure in subsection III.B.4 fuzzy weights are calculated. The ANN approach is shown to have the largest surface area for subgroup A, with a fuzzy weight of 0.84, followed by the multiple linear-regression method and moving average, with the fuzzy weights of 0.67 and 0.42, respectively.

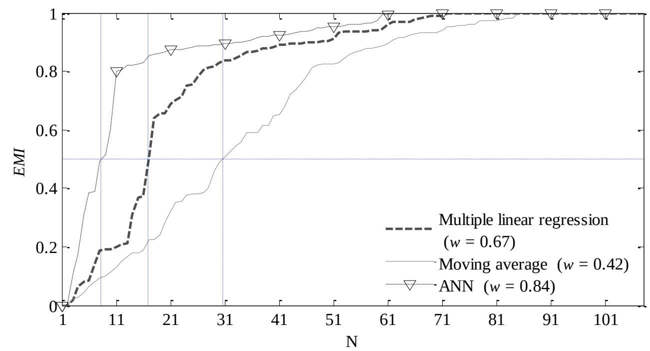


Fig. 9. Sorted membership function values *EMI* for historical absolute forecast errors for the chosen day, hour 16.

Fig. 10 shows the assigned fuzzy weight for each hour for the chosen forecasted day. The ANN method was shown to give the best results overall, followed by the multiple linear regression and the moving average.

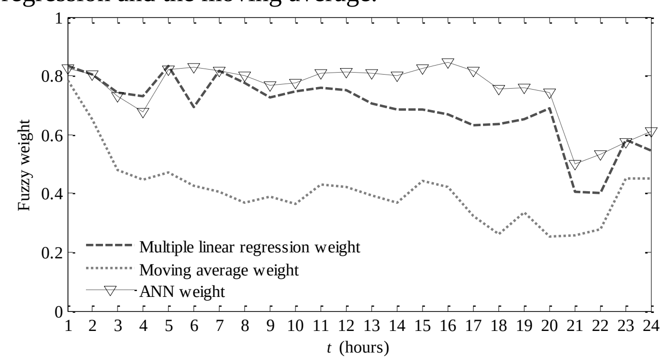


Fig. 10. Fuzzy weights for individual forecasting methods for the chosen day.

Fig. 11 shows an example of the day-ahead system generation forecasts for the Slovenian power system for the randomly chosen weekday and the MAPE of the forecasts for the day. The combined forecast has the lowest MAPE for the chosen day, with the MAPE of 4.2, compared to the MAPE of 9.06, 3.53 and 6.87 for multiple linear regression, the moving average and the ANN forecast, respectively.

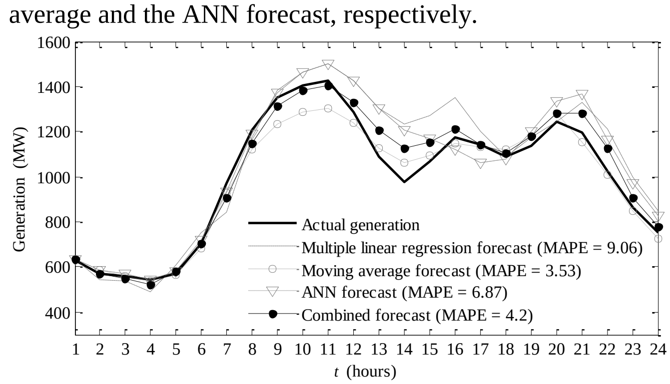


Fig. 11. Comparison of generation forecasting results for the chosen day for the Slovenian power system.

The network-topology forecasting is performed using information based on the transmission system operators' maintenance schedules and the day-to-day system-operation status, which is published on the system operator's homepage [20] and are considered to be always accurate. Table IV shows an example of the network-topology impact on active-power losses in cases of two different line switch-offs:

- 400 kV interconnection line Divaca-Redipuglia between Slovenia and Italy and
- 400 kV line Bericevo-Podlog, which serves as the main connection line between the eastern and western part of Slovenia.

The simulation of network-topology change was performed for hour 10 and normalized for the base case (all elements in the system were operational). The results show that network-topology can greatly influence losses and the effectiveness of taking network-topology forecasts into account, with the transmission losses being reduced by 16 % during Divaca-Redipuglia interconnection switch-off and increased by 46 % with Bericevo-Podlog line switch-off.

TABLE IV
INFLUENCE OF NETWORK-TOPOLOGY CHANGE ON TRANSMISSION LOSSES FOR A TEST CASE FOR THE SLOVENIAN POWER SYSTEM

	Active-power losses (p.u.)
Base case (all elements are operational)	1
Divaca-Redipuglia 400 kV switch-off	0.84
Bericevo-Podlog 400 kV switch-off	1.46

To analyze the accuracy of the proposed approach for active-power-loss forecasts, power-flow calculations were performed for individual forecasting methods as well as the combined forecasting approach. Table V shows the MAPE and standard deviation of the transmission-loss forecasts for the Slovenian power system for the analyzed month. Using individual forecast methods the transmission-loss forecasts MAPE ranged from 10.45 to 12.21 and the standard deviation of MAPE from 10.11 to 12.93. Using the proposed combined forecasts the transmission-loss forecast accuracy significantly improves, with the MAPE and standard deviation of the MAPE of 7.27 and 7.83.

TABLE V
MAPE AND STANDARD DEVIATION COMPARISON OF THE TRANSMISSION-LOSSES FORECAST RESULTS FOR THE ANALYZED MONTH FOR THE SLOVENIAN POWER SYSTEM

	MAPE	St. deviation (%)
Multiple linear regression	11.93	11.21
Moving average	12.21	12.93
ANN	10.45	10.11
Combined forecast	7.27	7.83

Fig. 12 shows the transmission-loss forecast for the Slovenian power system for the selected day using the individual forecasts only and the combined forecast. The transmission-loss forecast MAPE with the use of individual forecasts ranges from 7.83 to 12.8, while the use of the combined forecast gives us the MAPE of 5.1 for the selected day. The forecasted required energy to cover the losses is compared to the current official long-term method of energy purchases in Slovenia, i.e., a yearly purchase of a 15 MWh/h base load energy and 10 MWh/h energy for peak loads [20].

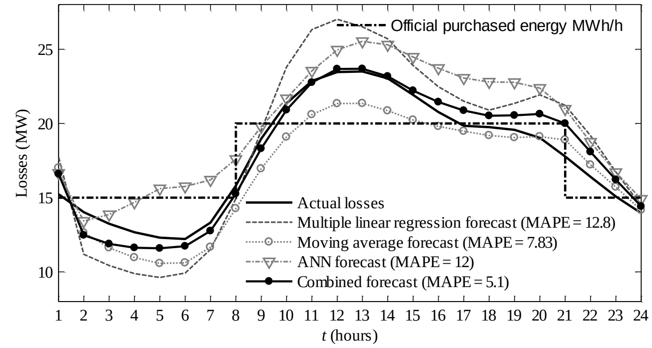


Fig. 12. Transmission losses for the Slovenian power system for the chosen day for the Slovenian power system.

B. *Case study with incomplete or missing data*

In case-study A, it is shown that the ANN method gives, on average, the best results. The question is, if that would also be the case when the data is partly missing. This is simulated in case-study B, where larger amounts of weather data are missing.

The load forecast results are presented in Fig. 13. The moving average gives better results than the multiple linear regression and the ANN in such cases, as it is insensitive to weather-data outages and represents a robust forecasting method.

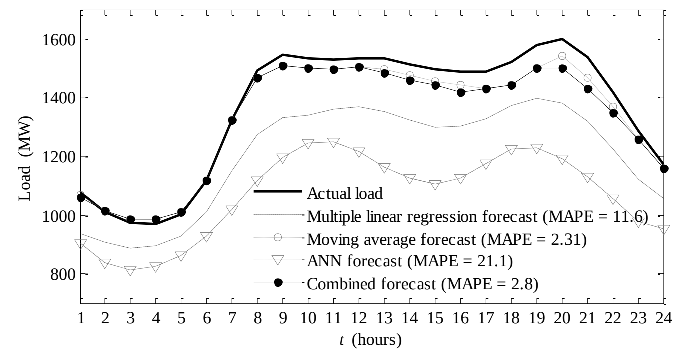


Fig. 13. Comparison of load-forecasting results for the chosen day for the Slovenian power system with missing or incomplete weather data.

The results show that the weather-sensitive methods give a high MAPE of 11.6 and 21.1 MAPE for the multiple linear regression and ANN, respectively, while the moving average approach has the MAPE of 2.31. The combined-result MAPE is slightly higher, as both the weather-sensitive methods must be taken into account at certain times due to the assigned fuzzy weights shown in Fig. 14.

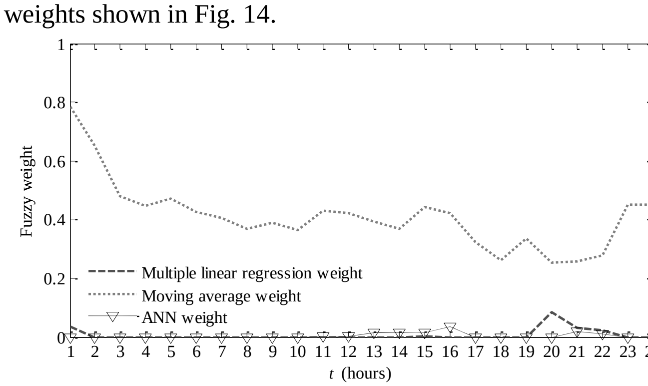


Fig. 14. Transmission losses for the Slovenian power system for the chosen day for the Slovenian power system.

Due to the clarity of the paper, only load forecasts results are shown below but similar findings are obtained for a generation forecast. Fig. 15 shows the transmission-loss forecast for the Slovenian power system for the selected day using the individual forecasts only and the combined forecast, where larger amounts of weather data are missing. The transmission-losses forecast MAPE with the use of individual weather-sensitive forecasts ranges up to 60.4, while the use of the combined forecast gives us the MAPE of 8 for the selected day, as the moving average is assigned high fuzzy weights, Fig. 14.

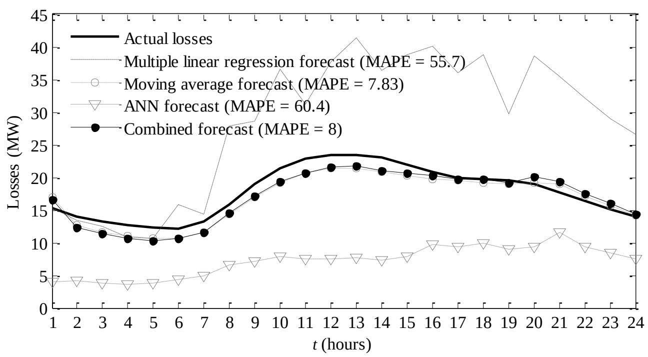


Fig. 15. Transmission losses for the Slovenian power system for the chosen day for the Slovenian power system where larger amounts of weather data are missing.

V. Conclusion

In this paper a novel, short-term, active-power-loss forecasting method is presented. The method is based on AC power-flow calculations of power system state-forecasts, i.e. load, generation and network topology. The system load and generation are forecasted for the entire interconnected power systems to take the power transits throughout the observed power system and system loading into account. Specifically, the weather-sensitive ANN, multiple linear regression approach and the non-weather-sensitive moving average are used to forecast the load and generation. To minimize the forecast errors and increase the robustness, a novel fuzzy-based grouping of the forecasting-method results is presented. The fuzzy weight-based combination of methods presents a novel way of rewarding or penalizing the accuracies and inaccuracies of individual forecasting methods.

The presented approach is currently being designed for the use in daily operations by the Slovenian system operator. The proposed forecasting method has been tested for a randomly chosen month for the Slovenian power system. According to the results the following conclusions can be drawn:

- the use of combined forecasts gives the lowest transmission-loss forecast MAPE and standard deviation, indicating improved accuracy compared to the use of individual forecasting methods;
- the use of combined forecasts represents a robust forecast method in the case of partly missing data, which can often occur in daily system operation;

The forecast results of the system load, generation and losses using the proposed approach verify its effectiveness and substantial financial savings are possible with the presented accuracy of the proposed approach.

References

[1] Sandrin, P., "System-services: several questions, in an integrated utility perspective [ancillary services]," *Pricing of Ancillary Services: an International Perspective (Digest No: 1996/164)*, IEE Colloquium on , vol., no., pp.7/1-7/5, 28 Jun 1996

[2] Swissgrid, Publication of the AS Concept on the Internet, Swissgrid, 2008.
Available: http://www.swissgrid.ch/power_market/grid_operation/

[3] Chow J. H.; Wu F. F.; Momoh J. A., *Applied Mathematics for Restructured Electric Power Systems*. New York: Springer US, 2006, ch. 12

[4] Moghaddas-Tafreshi, S.M.; Farhadi, M., "A linear regression-based study for temperature sensitivity analysis of Iran electrical load," *Industrial Technology, 2008. ICIT 2008. IEEE International Conference on* , vol., no., pp.1-7, 21-24 April 2008

[5] Ruzic, S.; Vuckovic, A.; Nikolic, N., "Weather sensitive method for short term load forecasting in Electric Power Utility of Serbia," *Power Systems, IEEE Transactions on* , vol.18, no.4, pp. 1581-1586, Nov. 2003

[6] Hagan, M. T.; Behr, S. M., "The Time Series Approach to Short Term Load Forecasting," *Power Systems, IEEE Transactions on* , vol.2, no.3, pp.785-791, Aug. 1987

[7] Hong-Tzer Y.; Chao-Ming H., "A new short-term load forecasting approach using self-organizing fuzzy ARMAX models," *Power Systems, IEEE Transactions on* , vol.13, no.1, pp.217-225, Feb 1998

[8] Taylor J.W., "Short-term electricity demand forecasting using double Seasonal exponential smoothing," *Journal of the Operational Research Society*, Volume 54, Number 8, August 2003 , pp. 799-805(7)

[9] Villalba, S.A.; Bel, C.A., "Hybrid demand model for load estimation and short term load forecasting in distribution electric systems," *Power Delivery, IEEE Transactions on* , vol.15, no.2, pp.764-769, Apr 2000

[10] Yang, J., Stenzel, J., "Short-term load forecasting with increment regression tree," *Electric Power Systems Research*, Volume 76, Issues 9-10, June 2006, Pages 880-888

[11] Jain, A.; Satish, B., "Clustering based Short Term Load Forecasting using Support Vector Machines," *PowerTech, 2009 IEEE Bucharest* , vol., no., pp.1-8, June 28 2009-July 2 2009

[12] Yang J., "Power System Short-Term Load Forecasting," Darmstadt, Fachbereich Elektrotechnik und Informationstechnik, 2005. Available: http://tuprints.ulb.tu-darmstadt.de/epda/000662/

[13] Peng, T.M.; Hubele, N.F.; Karady, G.G., "Advancement in the application of neural networks for short-term load forecasting," *Power Systems, IEEE Transactions on* , vol.7, no.1, pp.250-257, Feb 1992

[14] Murto P., "Neural Network Models for Short-Term Load Forecasting," Helsinki, Helsinki University of Technology.
Available: http://www.sal.hut.fi/Publications/pdf-files/tmur98.pdf

[15] Kiartzis, S.J.; Bakirtzis, A.G.; Theocharis, J.B.; Tsagas, G., "A fuzzy expert system for peak load forecasting. Application to the Greek power system," *Electrotechnical Conference, 2000. MELECON 2000. 10th Mediterranean* , vol.3, no., pp. 1097-1100 vol.3, 29-31 May 2000

[16] Kwang-Ho K.; Hyoung-Sun Y.; Yong-Cheol K., "Short-term load forecasting for special days in anomalous load conditions using neural networks and fuzzy inference method," *Power Systems, IEEE Transactions on* , vol.15, no.2, pp.559-565, May 2000

[17] Klir G.J.; Yaun B., Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice Hall PTR, May 1995

[18] Yuan-Sheng H., Jia-Jia Deng, "Short-Term Load Forecasting Based on Ant Colony Fuzzy Clustering and SVM Algorithm," Natural Computation, 2008. ICNC '08. Fourth International Conference on , vol.2, no., pp.162-166, 18-20 Oct. 2008

[19] Hu G., Zhu F.F., Zhang Y.Z., "Short-Term Load Forecasting Based on Fuzzy C-Mean Clustering and Weighted Support Vector Machines," International Conference on Natural Computation, vol. 5, pp. 654-659, Third International Conference on Natural Computation, 2007

[20] ELES, Rules for the purchase of energy to cover losses on the transmission system in the year 2010. Available: http://www.eles.si/en/for-business-users/system-services.aspx

BIOGRAPHIES

Matej Rejc received his Diploma Engineer degree from the University of Ljubljana, Slovenia, in 2007. He is employed as an assistant at the University of Ljubljana, Faculty of Electrical Engineering. His main research field includes power-system operation, protection and control and ancillary services.

Miloš Pantoš was born in 1977 in Slovenia. He received his Diploma Engineer degree and Dr. Sc. degree from the University of Ljubljana, Slovenia, in 2001 and 2005, respectively. He is employed as an associate professor at the University of Ljubljana, Faculty of Electrical Engineering. He is a head of the Laboratory of Power Systems. His main research field includes power-system operation, protection and control, power market and ancillary services.