International Journal of Innovative Computing, Information and Control
Volume **10**, Number **4**, August **2014**
ICIC International ©2014 ISSN 1349-4198
pp. 1435–1445

SMOOTH COMPONENT ANALYSIS AND MSE DECOMPOSITION FOR ENSEMBLE METHODS IN MULTI-AGENT ENVIRONMENT

Ryszard Szupiluk[1], Piotr Wojewnik[1] and Tomasz Zabkowski[2]

[1]Collegium of Economic Analysis
Warsaw School of Economics
Al. Niepodleglosci 162, 02-554 Warsaw, Poland
rszupi@sgh.waw.pl; piotr.wojewnik@gmail.com

[2]Faculty of Applied Informatics and Mathematics
Warsaw University of Life Sciences
Nowoursynowska 159, 02-787 Warsaw, Poland
tomasz_zabkowski@sggw.pl

Received December 2012; revised June 2013

Abstract. *This paper is addressed to economic problems for which many different solutions (models) can be proposed. In such situation the ensemble approach is a natural way to improve the final prediction results. In particular, we present the method for the prediction improvement in multi-agent environment based on the multivariate decompositions. As a decomposition method we present the smooth component analysis. The resulting components are classified as destructive and then removed or as constructive and then recomposed to create final forecast. The classification of the components is based on the theoretical analysis of MSE error measure. The robustness of the method was validated based on energy load data from Polish power system.*
Keywords: Ensemble methods, Multivariate decompositions, Smooth component analysis, Agent based modeling

1. **Introduction.** The term model plays an important role in sciences, but the meaning and the scope of modeling depends on the specific application areas, schools or even the authors. The common feature linking these different approaches towards model definition is the awareness that the model presents only a given perspective of the real phenomenon and it is considered only as useful analytical representation of the given problem. In this sense we can find out that each real process or phenomenon may be represented by a few different models. It seems that one of the fundamental challenges facing both academics and practitioners is to develop an effective method to combine information coming from different theoretical approaches. In particular, this refers to the quantitative results. The advantages of models aggregation are particularly evident in case of agent systems, machine learning or data mining models, what in fact involves all approaches that are based on mathematical methods of data exploration [8,16]. Let us bear in mind that although the ensemble methods are currently popular research area, the aggregation of different outcomes with a completely different methodological approaches is not obvious. Most of the existing aggregation methods such as boosting, bagging, or stacked regression require quite restrictive assumptions about the parameters, and the structure of the results and variables distributions as well [3,9]. In most cases, aggregation concerns models of the same structure or even the same models but estimated on other subsets of the training data.

In this paper we develop the approach based on blind signal separation methods in a multi-agent environment. In this concept, we focus rather on the physical meaning and interpretation of the data (variables) than on their formal mathematical properties. We assume that the results generated by different intelligent agents possess certain physical components that disrupt the prediction. These components may be related to the inaccuracy, inadequacy or noisy input data, but they can also be the result of improper models specification or wrong learning algorithm choice. As a result, we can assume that the results are the combination of certain constructive and, from the other side, destructive components which are responsible for the prediction errors. Therefore, separation and then elimination of these destructive components should bring prediction improvement [22].

For this purpose we apply smooth component analysis to find the latent components [22] and in the next step we classify them as destructive or constructive. For this task we propose analysis of second order statistics with mean square error criterion (MSE). The validation of this approach is performed based on practical experiment with energy load prediction from Polish market [13].

2. **Agent Based Modeling for Prediction Improvement.** Agent-based systems have attracted much attention in recent years because of its promise as a new paradigm for designing, creating and implementing systems that operate in distributed, complex or very dynamic environments. The basic units of an agent based systems are agents, which are considered to be autonomous and adaptive. That means there is no control over their behavior and they can react to changes in the environment. Each agent in a multi-agent system represents a specific set of problem solving skills and experience and as a whole they perform better for a given problem solving. The idea of agent is better realized within the framework of complex adaptive systems (CAS) [2]. CAS is a complex system of interacting units, which include goal-directed units, that is, units that are reactive and that direct at least some of their reactions towards the achievement of built-in or evolved goals [23]. This broad definition enables agents to be entities ranging from active data-gathering, and decision-making with sophisticated learning algorithms to passive units with no cognitive functions assigned. A modern market oriented economy is an example of a CAS, consisting of a collection of autonomous agents interacting in various market contexts. In this stream an agent based modeling (or agent-based computational economics) can be distinguished [18,23] which is the computational approach to study economic processes as dynamic systems consisting of interacting agents. Therefore, this concept has motivated researchers to adopt it for the study of several electricity market issues. Some authors have concerned agent based models for examining electricity consumer behavior at the retail level, e.g., [17,19] or for analyzing distributed generation models, e.g., [12,20].

In this paper, we focus on an application of the ensemble approach (information aggregation) to prediction improvement. The approach is especially suited for complex and dynamic problems such as energy load prediction.

The problem of information aggregation and synthesis gathered from data sources arises in artificial intelligence very often. An example may be an artificial expert system that is responsible to conclude from the knowledge of human experts on a specific topic (a disease diagnosis, a portfolio choice, weather prediction, etc.). The pieces of information provided by experts are represented in a formal language and the point is to merge them into standardized database of unique knowledge. The artificial agents are driven by the concept of taking the best possible decisions and therefore they can act as the information sources or information processing units. The recent field of information fusion concerns how to aggregate individual information into a collective one. Therefore, we will consider how to combine the knowledge of several intelligent agents with application of smooth component analysis and MSE decomposition. In particular, we focus on short term load forecasting which plays an important role in the formulation of economic, reliable and secure strategies for the power system. The proposed prediction improvement approach is defined in multi-agent framework, in which we can distinguish various types of intelligent agents to perform different tasks [1].

The *data retrieval agents* (or data mining agents) are the units to communicate with source data and perform data retrieval and data pre-processing taking into account the algorithms used later on. Based on the data specification obtained from the user the data retrieval agents will maintain its interaction with the sources to collect all the data related to historic load and any additional information, e.g., weather forecasts. In general, this kind of agents will be responsible for the preprocessing of the most relevant data for the training and testing of artificial neuron networks (ANNs), fuzzy logic techniques or expert systems.

Intelligent agents. This kind of agents will use for instance, ANNs, knowledge based systems or fuzzy logic techniques to perform the following tasks:

(1) To generate a set of ANNs trained over different time windows;
(2) To aggregate the information from the ANN models using smooth component analysis and MSE error decomposition.

This system will do the same reasoning that is provided by intuitive forecasting of electrical load but it will be reduced to a couple of formal steps. In this approach, a variety of well-tuned models will be available in the form of models library and be accessed by the system to make the most relevant forecast. In order to make the forecasting system more robust to change in the environment, we use multiple agents for final forecast. Therefore, we distinguish N adaptive forecasting agents called intelligent agents which are represented by diverse multilayer perceptron (MLP) artificial neural networks. Such agent oriented architecture can result in forecast improvement. In particular, this is the focus of this paper.

Interface Agent. This adopted approach requires also an additional interface agent for communication with the other agents and the end users. It allows activities such as presenting the output or reporting differences between forecasts and the actual values. The final answer will be synthesized based on integrating the resultant outputs acquired from intelligent agents for an overall solution for load prediction problem. The information sharing among different intelligent agents allows the system to produce a better forecast.

In this work, we use agent-based modeling approach as a computational method that enables to create, analyze, and experiment with models composed of agents that interact within the environment [18]. Therefore, proposed prediction approach is defined in multi-agent framework with no distinction to specific agent platform or architecture. It is the specific problem which determines the structure of the environment for agents' application and affects the way in which agents interact or communicate.

3. **The Framework for Ensemble Method.**

In this paper, we assume that after learning various intelligent agents represented by neural network models we have a set of prediction results. For simplicity, in further consideration we assume that our models results x_i and target p are centered. In practice, it means that before models integration we remove mean values from prediction results and target, and after integration we add this values.

We collect particular prediction results x_i, $i = 1, \dots, m$, in one multivariate variable $\boldsymbol{x}(k) = [x_1(k), \dots, x_m(k)]^T$. Now we assume that prediction results $x_i(k)$ is a mixture of the latent components: constructive $\hat{s}_j(k)$ is associated with the predicted variable, and destructive $s_j(k)$ is associated with the inaccurate and missing data, imprecise estimation, distribution assumptions, etc. We assume the relation between observed prediction results and latent components to be represented as

$$\boldsymbol{x}(k) = \boldsymbol{A}\boldsymbol{s}(k), \tag{1}$$

where $\boldsymbol{s}(k) = [\hat{s}_1(k), \ldots, \hat{s}_i(k), s_{i+1}(k), \ldots, s_m(k)]^T$, and matrix $\boldsymbol{A} \in R^{m\times m}$ represents the mixing system. The relation (1) stands for decomposition of prediction results x_i into latent components matrix $\mathbf{s}$ and mixing matrix $\boldsymbol{A}$. Our aim is to find $\boldsymbol{s}$ and $\boldsymbol{A}$, reject its destructive part (replace signals $s_j(k)$ with zero) and next to mix the constructive components back to obtain improved prediction results:

$$\hat{\boldsymbol{x}}(k) = \boldsymbol{A}\hat{\boldsymbol{s}}(k) = \boldsymbol{A}[\hat{s}_1(k), \ldots, \hat{s}_i(k), 0_{i+1}, \ldots, 0_m]^T. \tag{2}$$

The crucial point of the above concept is proper $\boldsymbol{A}$ and $\boldsymbol{s}$ estimation. This problem can be described as Blind Signal Separation task [5,10,20] which aims to find such matrix $\boldsymbol{W} = \boldsymbol{A}^{-1}$ that

$$\boldsymbol{s}(k) = \boldsymbol{W}\boldsymbol{x}(k). \tag{3}$$

The BSS methods explore different properties of data like: independence [4,5,10], decorrelation [5], sparsity [15], smoothness [5,22], and non-negativity [14]. In this paper, we focus on Smooth Component Analysis (SmCA) what is adequate for data with temporal structure. Smooth component analysis is a method for the smooth components identification in a multivariate variable [5,22]. For N-observation signals with temporal structure we propose a following smoothness measure

$$P(s) = \frac{\frac{1}{N}\sum\limits_{k=2}^{N} |s(k) - s(k-1)|}{\max(s) - \min(s) + \delta(\max(s) - \min(s))}, \tag{4}$$

where symbol $\delta(.)$ means zero indicator function – valued at 0 everywhere except 0, where the value of $\delta(.)$ is 1. Measure (4) has straightforward interpretation: it is maximal when the changes in each step are equal to range (maximal possible change during one period), and is minimal when data are constant. The possible values vary from 0 to 1. Zero indicator $\delta(.)$ term is introduced to avoid dividing by zero.

The components are taken as linear combination of signals x_i and should be as smooth as possible. Our aim is to find such matrix $\boldsymbol{W} = [w_1, w_2, \ldots, w_n]$ that for $\boldsymbol{s} = \boldsymbol{W}\boldsymbol{x}$ we obtain $\boldsymbol{s} = [s_1, s_2, \ldots, s_m]^T$ where s_1 maximizes $P_1(s_1)$ so we can write

$$\boldsymbol{w}_1 = \underset{||\boldsymbol{w}||=1}{\arg\max}(P(\boldsymbol{w}^T\boldsymbol{x})). \tag{5}$$

Having estimated the first $k-1$ smooth components, the next one is calculated as least smooth component of the residual obtained in Gram-Schmidt orthogonalization [7]:

$$\boldsymbol{w}_k = \underset{||\boldsymbol{w}||=1}{\arg\max}\left(P\left(\boldsymbol{w}^T\left(\boldsymbol{x} - \sum_{i=1}^{k-1}\boldsymbol{s}_i\boldsymbol{s}_i^T\boldsymbol{x}\right)\right)\right), \tag{6}$$

where $\boldsymbol{s}_i = \boldsymbol{w}_i^T\boldsymbol{x},\ i = 1\ldots k$. As the numerical algorithm for finding $\boldsymbol{w}_n$ we can employ the conjugate gradient method with golden search as a line search routine. The algorithm outline for initial $\boldsymbol{w}_i(0) = rand$, $\boldsymbol{p}_i(0) = -\boldsymbol{g}_i(0)$ is as follows:

1. Identify the indexes l for extreme signal values:

$$\boldsymbol{w}_k = \underset{||\boldsymbol{w}||=1}{\arg\max}\left(P\left(\boldsymbol{w}^T\left(\boldsymbol{x} - \sum_{i=1}^{k-1}\boldsymbol{s}_i\boldsymbol{s}_i^T\boldsymbol{x}\right)\right)\right), \tag{7}$$

$$l^{\max} = \underset{l\in 1\ldots N}{\arg\max}\, \boldsymbol{w}_i^T(k)\boldsymbol{x}(l), \tag{8}$$

$$l^{\min} = \arg\min_{l \in 1 \ldots N} \boldsymbol{w}_i^T(k)\boldsymbol{x}(l). \tag{9}$$

2. Calculate gradient of $P(\boldsymbol{w}_i^T\boldsymbol{x})$:

$$\boldsymbol{g}_i = \frac{\partial P(\boldsymbol{w}_i^T\boldsymbol{x})}{\partial \boldsymbol{w}_i} = \frac{\sum\limits_{l=2}^{N} \Delta\boldsymbol{x}(l) \cdot sign(\boldsymbol{w}_i^T\Delta\boldsymbol{x}(l)) - P(\boldsymbol{w}_i^T\boldsymbol{x}) \cdot (\boldsymbol{x}(l^{\max}) - \boldsymbol{x}(l^{\min}))}{\max(\boldsymbol{w}_i^T\boldsymbol{x}) - \min(\boldsymbol{w}_i^T\boldsymbol{x}) + \delta(\max(\boldsymbol{w}_i^T\boldsymbol{x}) - \min(\boldsymbol{w}_i^T\boldsymbol{x}))}, \tag{10}$$

where $\Delta\boldsymbol{x}(l) = \boldsymbol{x}(l) - \boldsymbol{x}(l-1)$.

3. Identify the search direction (Polak-Ribiere formula [7])

$$\boldsymbol{p}_i(k) = -\boldsymbol{g}_i(k) + \frac{\boldsymbol{g}_i^T(k)(\boldsymbol{g}_i(k) - \boldsymbol{g}_i(k-1))}{\boldsymbol{g}_i^T(k-1)\boldsymbol{g}_i(k-1)}\boldsymbol{p}_i(k-1), \tag{11}$$

and calculate the new weights:

$$\boldsymbol{w}_i(k+1) = \boldsymbol{w}_i(k) + \alpha(k) \cdot \boldsymbol{p}_i(k), \tag{12}$$

where $\alpha(k)$ is found in golden search.

The above optimization algorithm should be applied as a multi-start technique with random initialization.

The orthogonalization process (6) after identification of each component ensures that components are ordered by their smoothness with correlation matrix given by:

$$\boldsymbol{R}_{ss} = E\left\{\boldsymbol{s}\boldsymbol{s}^T\right\} = \boldsymbol{D}, \tag{13}$$

where $\boldsymbol{D}$ is a diagonal matrix. The property (7) will be explored for destructive component identification where prediction is scored by MSE criterion.

4. **Extreme Value Distribution Preprocessing for SmCA.** The main disadvantage of the measure (4) is its high sensitivity to the outliers due to minimum and maximum value in denominator. To avoid this problem and to make the measure more robust, we can apply the estimation based on the generalized extreme value distribution (GEVD).

In this section, we present the basic properties of generalized extreme value distribution GEVD and our contribution to the smoothness measurement. We show that the smoothness value estimated directly from a signal is less effective, than estimated from the signal regularized by the extreme value distribution. We show also that in some cases the regularization influences the smoothness measure much more than simple scaling. There are signals, e.g., heavy tailed, with high probability that each particular observation will change the extremes a lot, and therefore the smoothness measure (4), too. It would be possible to stabilize the measure (4), using not necessarily empirical but rather representative extreme values. Therefore, we propose to estimate the representative extremes using the extreme value distribution [6,11]. The probability density function $f(z)$ of the generalized extreme value distribution with the location parameter μ, the scale parameter σ, and the shape parameter $\gamma \neq 0$ is represented by

$$f(z) = \frac{1}{\sigma}\left(1 + \gamma\frac{z-\mu}{\sigma}\right)^{-1-\frac{1}{\gamma}} \exp\left(-\left(1 + \gamma\frac{z-\mu}{\sigma}\right)^{-1/\gamma}\right), \tag{14}$$

for $1 + \gamma\frac{z-\mu}{\sigma} > 0$, where $\gamma > 0$ (Type II) or $\gamma < 0$ (Type III). For $\gamma = 0$ (Type I) GEVD is

$$f(z) = \frac{1}{\sigma}\exp\left(-\exp\left(\frac{z-\mu}{\sigma}\right) - \frac{z-\mu}{\sigma}\right). \tag{15}$$

From the method of moments we can estimate

$$\mu = \frac{\bar{\sigma}\sqrt{6}}{\pi} \tag{16}$$

and

$$\sigma = \bar{z} - 0.5772\mu \tag{17}$$

where $\bar{z}$ and $\bar{\sigma}$ are the sample mean and standard deviation, respectively.

Consequently, we can propose the algorithm for smoothness estimation in signal s_i^α which is our main contribution to the GEVD approach, defined as follows:

1. From the signal, generate the bootstrap samples and calculate their min's z;
2. Assume z's are realizations of $f(z)$ distribution and fit GEVD with μ_- as the location parameter for minimums;
3. For maximums calculate the location parameter μ_+ of $f(-z)$, respectively;
4. Regularize the signal s_i^α to $s_{reg_i}^\alpha$ by correction of outliers to range $[\mu_-, \mu_+]$;
5. Calculate the smoothness measure $P(s)$ for regularized signal.

5. **Destructive Components Identification.** The MSE is one of the most popular criteria for model scoring. In our case we can describe MSE_i for each x_i as

$$MSE_i = E\{(p - x_i)^2\}, \tag{18}$$

where p is target variable. According to our assumptions the x_i can be expressed as linear combination of the latent components, which leads us to

$$MSE_i = E\left\{\left(p - \sum_{j=1}^{m} a_{ij}s_j\right)^2\right\} = E\left\{p^2 + \sum_{j=1}^{m}\sum_{l=1}^{m} a_{ij}a_{il}s_js_l - 2p\sum_{j=1}^{m} a_{ij}s_j\right\}. \tag{19}$$

After SmCA the latent components are decorrelated, that is, $E\{s_is_j\} = 0$, so

$$MSE_i = \sigma_p^2 + \sum_{j=1}^{m}\left(a_{ij}^2\sigma_{s_j}^2 - 2a_{ij}\rho_{p,s_j}\right), \tag{20}$$

where $\sigma_p^2 = E\{p^2\}$, $\sigma_{s_j}^2 = E\{s^2\}$ and $\rho_{p,s_j} = E\{ps_j\}$.

Presented MSE_i calculation for x_i prediction explains what happens after elimination of $a_{ij}s_j$ from x_i for every $j = 1, \dots, m$. Namely, if the condition

$$\sigma_{s_j}^2 \le \frac{2\rho_{p,s_j}}{a_{ij}}, \tag{21}$$

holds, we can expect reduction of MSE_i by value

$$\vartheta_{ij} = a_{ij}^2\sigma_{s_j}^2 - 2a_{ij}\rho_{p,s_j}. \tag{22}$$

Therefore, the value ϑ_{ij} can be used for choosing the component s_j that is responsible for the highest MSE_i reduction.

The whole framework, starting from data retrieval and including agents modeling phase with smooth components analysis and destructive components identification is presented in Figure 1.

Here, we can distinguish various types of intelligent agents performing different tasks in proposed approach, that is, data retrieval agents, interface agent and in particular, intelligent agents. The latter ones are the subject of our approach, thus generating robust in dynamic environment to produce a better load forecast. In our case, proposed multi-agent environment may be viewed as the box with a group of intelligent agents inside. The subsequent elements of the input sequences are supplied to the environment, where they become available for all agents. Each agent itself may analyze the incoming data and produce individual predictions and for each agent an artificial neural network may be used as a basic mechanism to trace signal regularities for time series prediction.

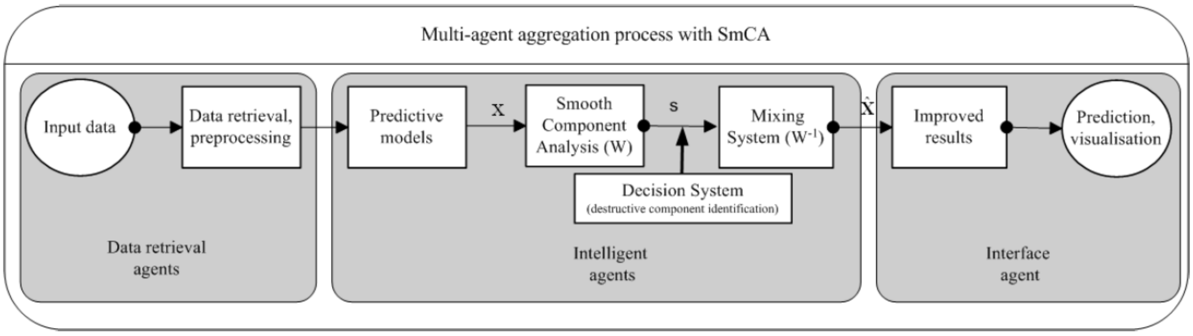

FIGURE 1. Multi-agent aggregation process with SmCA and destructive components removal

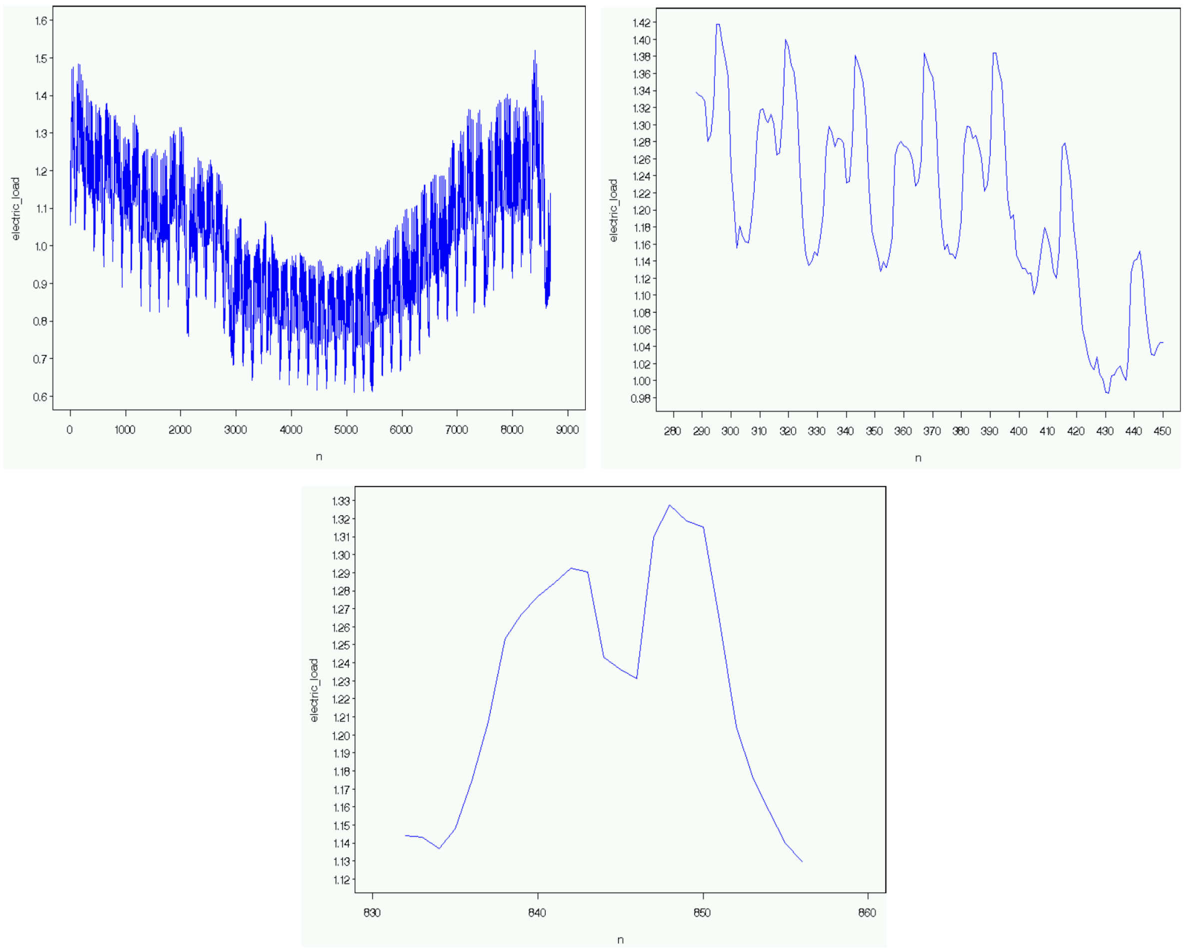

FIGURE 2. Yearly, weekly and daily load data covering year 1997 (vertical axis shows normalized load and horizontal axis shows subsequent observations)

6. **Practical Experiment.** To verify the validity of the concept, we used the data from Polish power system. The data set included 86400 observations (hourly data) covering time span of 1988-1998 years. Figure 2 presents an hourly load demand for the year 1997, a given week of 1997 and a given day of 1997.

In agent-based modeling approach the system consists of a set of intelligent agents that encapsulate the behaviors of various individuals that make up the system. In our paper, intelligent modeling agents are made using artificial intelligence properties thus possessing some aspects which are capable of flexible autonomous action to meet their objectives. Therefore, we distinguish N adaptive forecasting agents called intelligent agents which are represented by diverse multilayer perceptron (MLP) artificial neural networks. In particular, we build six neural networks with different learning methods (delta, quasi-Newton, Levenberg-Marquard) to forecast hourly energy consumption in Poland in next 24 hours. The available variables to create the forecast included energy demand from the last 24 hours and calendar variables such as month, day of the month, day of the week, and holiday indicator. The Test1 data set (43200 observations) is used to estimate the decomposition matrices $\boldsymbol{W}$, $\boldsymbol{A}$ and to calculate the expected values of MSE reduction ϑ_{ij}. In the final phase we use Test2 data set (43200 observations excluded from the previous analysis) to calculate the MSE reduction obtained after physical elimination of particular components.

The quality of the neural models on the Test1 data set is presented in Table 1.

The smooth components analysis applied to the prediction results from the Test1 data gives $\boldsymbol{A}$, $\boldsymbol{W}$ and the components presented in Figure 3.

The smoothness $P(s_j)$ for identified signals is presented in Table 2.

According to (12) for each model x_i, $i = 1, \ldots, 6$ and for each signal s_j, $j = 1, \ldots, 6$, we calculated expected level of MSE reduction ϑ_{ij}, see Table 3. In particular, the component s_1 is constructive, because its elimination would increase the MSE of each model. The component s_5 is destructive; therefore, we conclude that its elimination would decrease

Table 1. MSE of the neural models on the Test1 data set

MLP, Learning	28:11:1, Quasi-Newton	28:11:1, Delta	28:10:1, Levenberg-Marquard	28:11:1, Levenberg-Marquard	28:12:1, Levenberg-Marquard	28:13:1, Levenberg-Marquard
MSE [$\times 10^5$]	5,76	5,96	6,55	6,52	6,46	6,40

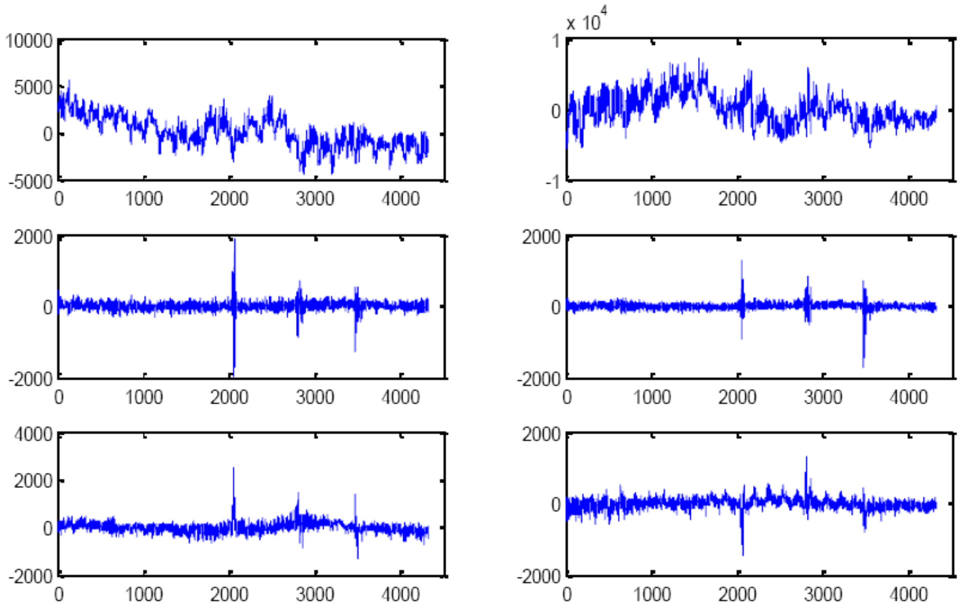

Figure 3. Smooth components obtained on the Test1 data (vertical axis shows the value of the components and horizontal axis shows subsequent observations)

Table 2. Signal smoothness $P(s_j)$ calculated on the Test1 data set

Component	s_1	s_2	s_3	s_4	s_5	s_6
$P(s_j)$	0,0287	0,0293	0,0363	0,0371	0,0405	0,0649

Table 3. Expected MSE reduction ϑ_{ij} calculated on the Test1 data set. Positive values denote value of MSE reduction possible to be obtained after elimination of s_j from x_i.

ϑ_{ij} [$\times 10^5$]	28:11:1, Quasi-Newton	28:11:1, Delta	28:10:1, Levenberg-Marquard	28:11:1, Levenberg-Marquard	28:12:1, Levenberg-Marquard	28:13:1, Levenberg-Marquard
s_1	−58,19	−58,47	−58,33	−58,31	−58,36	−58,44
s_2	−9,72	−9,15	−8,04	−8,08	−8,07	−8,09
s_3	0,05	0,1	0	0	0,05	0,08
s_4	0,04	0,08	0,06	0	0,03	0
s_5	0,82	0,45	0	0	0	0,04
s_6	0,02	0,13	0,03	0,03	0,03	0

Table 4. Quality of the primary models on the Test2 data set

MLP, Learning	28:11:1, Quasi-Newton	28:11:1, Delta	28:10:1, Levenberg-Marquard	28:11:1, Levenberg-Marquard	28:12:1, Levenberg-Marquard	28:13:1, Levenberg-Marquard
MSE [$\times 10^5$]	7,61	8,08	5,33	5,22	5,75	5,75

Table 5. Positive values denote MSE reduction obtained after physical elimination of component s_j from prediction x_i on the Test2 data set

ΔMSE_{ij} [$\times 10^5$]	28:11:1, Quasi-Newton	28:11:1, Delta	28:10:1, Levenberg-Marquard	28:11:1, Levenberg-Marquard	28:12:1, Levenberg-Marquard	28:13:1, Levenberg-Marquard
s_1	−58,66	−64,25	−59,78	−59,2	−60,49	−61,2
s_2	−6,97	−9,13	−10,62	−10,67	−10,24	−10,11
s_3	0,22	0,43	−0,03	0,03	−0,1	0,29
s_4	0,47	0,76	−0,23	0,13	0,36	0,07
s_5	1,83	1,44	−0,02	0,01	−0,13	−0,26
s_6	−0,07	−0,49	−0,05	−0,05	−0,07	−0,01

the MSE of each model. In this case, what is worth mentioning, rejection of component s_5 reduces the MSE error by 14.2% ((0,82×10^5)/(5,76×10^5) $= 0,142$) for agent represented by MLP 28:11:1 Quasi-Newton model, and by 7,8% ((0,45×10^5)/(5,76×10^5) $= 0,078$) for agent represented by MLP 28:11:1 Delta.

In Table 4, we present the quality of the models on the Test2 data.

We decomposed Test2 data set using $\boldsymbol{W}$ matrix estimated on Test1 data set. Then we physically eliminated the components and mixed the signals back using $\boldsymbol{A}$. In Table 5, we can observe obtained MSE reduction (ΔMSE_{ij}). In this case, rejection of component s_5 reduced the MSE error by 24.1% ((1,83×10^5)/(7,61×10^5) $= 0,241$) for agent represented by MLP 28:11:1 Quasi-Newton model, and by 18,9% ((1,44×10^5)/(8,08×10^5) $= 0,189$)

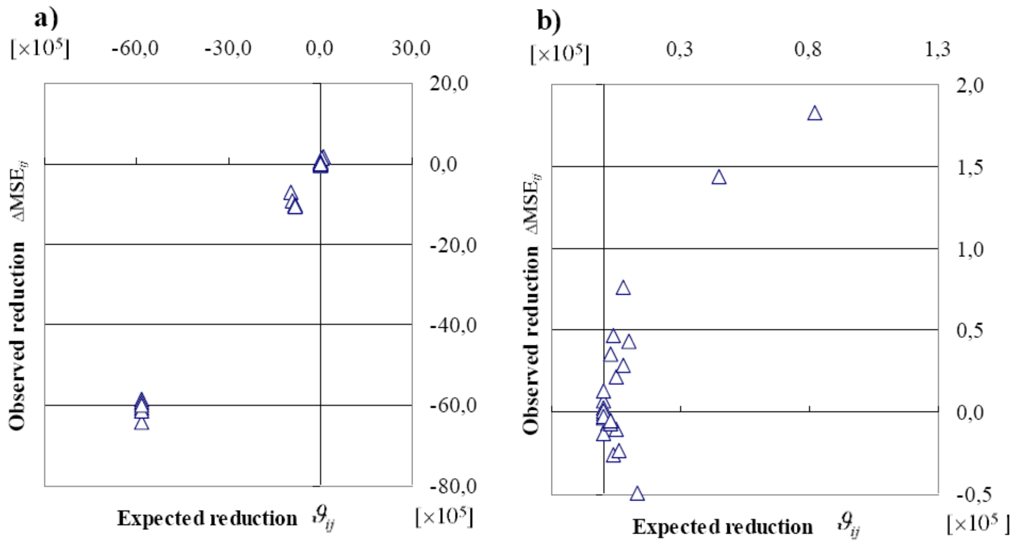

FIGURE 4. Expected reduction ϑ_{ij} versus obtained values ΔMSE_{ij}: a) for every (x_i, s_i) combination, b) for the case with positive ϑ_{ij} values

for agent represented by MLP 28:11:1 Delta. We also reported that in case of particular models we could observe that rejection of a given component resulted in worsening the prediction. In particular, this was the case with MLP using nonlinear Levenberg-Marquard algorithm for optimization. Such primary models prepared on Test2 data were quite well tuned, that is, they gave quite low MSE errors even before any postprocessing.

Next, in Figure 4 we present observed ΔMSE_{ij} in comparison to the expected ϑ_{ij}. This practical experiment proved validity of the concept of MSE reduction. In particular, we observed correlation between ϑ_{ij} and ΔMSE_{ij} at the level of 0.99.

7. **Conclusions.** In this paper, we considered the integration of the information generated by different intelligent modeling agents using smooth components analysis. For the MSE criterion we presented the theoretical background for efficient classification of the latent components. The results from the experiment confirmed the rationality of the approach.

We mainly focused on the results decomposition based on smooth component analysis, but the above identification method can be addressed to wide area of data exploration models, including simulations or machine learning systems. In particular, presented approach can be applied in trading systems, where the techniques that can automatically identify the fundamental determinants of the stock market are needed. Unfortunately, these factors are often hidden or mixed with noises. Therefore, a fundamental problem in financial market modeling is to estimate the main trends and to separate the general market dependencies from the individual behavior of a given financial instrument. This leads directly to the issue of data decomposition and interpretation of the underlying hidden components which correspond to the research presented in this paper.

Acknowledgment. This work was supported by the National Science Center in Poland based on decision number DEC-2011/03/B/HS4/05092.

REFERENCES

[1] S. R. Asar, A. U. Hassnain and A. Khattack, Multi-agent approach to short term load forecasting problem, *International Journal of Intelligent Control and Systems*, vol.10, no.1, pp.52-59, 2005.

[2] L. Blume and D. Easley, Evolution and market behavior, *J. of Economic Theory*, vol.58, pp.9-40, 1992.
[3] L. Breiman, Bagging predictors, *Machine Learning*, vol.24, pp.123-140, 1996.
[4] J. F. Cardoso, High-order contrasts for independent component analysis, *Neural Computation*, vol.11, pp.157-192, 1999.
[5] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*, John Wiley, Chichester, 2002.
[6] M. Evans, N. Hastings and B. Peacock, *Statistical Distributions*, John Wiley and Sons, 2000.
[7] G. H. Golub and C. F. Van-Loan, *Matrix Computations*, Johns Hopkins, 1996.
[8] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Macmillan, New York, 1994.
[9] J. Hoeting, D. Modigan, A. Raftery and C. Volinsky, Bayesian model averaging: A tutorial, *Statistical Science*, vol.14, pp.382-417, 1999.
[10] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley, 2001.
[11] N. L. Johnson, S. Kotz and A. W. Kemp, *Univariate Discrete Distributions*, John Wiley and Sons, 1992.
[12] K. Kok, Z. Derzsi, J. Gordijn, M. Hommelberg, C. Warmer, R. Kamphuis and H. Akkerman, Agent-based electricity balancing with distributed energy resources, a multiperspective case study, *Proc. of the 41st Annual Hawaii International Conference on System Sciences*, Washington, DC, USA, 2008.
[13] A. Lendasse, M. Cottrell, V. Wertz and M. Verdleysen, Prediction of electric load using Kohonen maps – Application to the polish electricity consumption, *Proc. of American Control Conference*, Anchorage, USA, pp.3684-3689, 2002.
[14] D. D. Lee and H. S. Seung, Learning of the parts of objects by non-negative matrix factorization, *Nature*, vol.401, pp.788-791, 1999.
[15] Y. Li, A. Cichocki and S. Amari, Sparse component analysis for blind source separation with less sensors than sources, *The 4th Int. Symposium on ICA and Blind Signal Separation*, Japan, pp.89-94, 2003.
[16] T. Mitchell, *Machine Learning*, McGraw-Hill, Boston, 1997.
[17] M. Muller, F. Sensfuß and M. Wietschel, Simulation of current pricing-tendencies in the German electricity market for private consumption, *Energy Policy*, vol.35, no.8, pp.4283-4294, 2007.
[18] M. J. North and C. M. Macal, *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation*, Oxford University Press, 2007.
[19] J. Roop and E. Fathelrahman, Modeling electricity contract choice: An agent-based approach, *Proc. of the ACEEE Summer Study Meeting*, Rye Brook, New York, USA, 2003.
[20] S. Rumley, E. Kagi, H. Rudnick and A. Germond, Multi-agent approach to electrical distribution networks control, *Proc. of the 32nd Annual IEEE International Computer Software and Applications Conf.*, Washington, DC, USA, pp.575-580, 2008.
[21] J. V. Stone, Blind source separation using temporal predictability, *Neural Computation*, vol.13, no.7, pp.1559-1574, 2001.
[22] R. Szupiluk, P. Wojewnik and T. Zabkowski, Smooth component analysis as ensemble method for prediction improvement, *Lecture Notes in Computer Science*, vol.4666, pp.277-284, 2007.
[23] L. Tesfatsion and K. Judd, Handbook of computational economics: Agent-based computational economics, *Handbook in Economics Series*, vol.2, 2006.