Evolving Smart Meter Data Driven Model for Short-Term Forecasting of Electric Loads

Harri Niska
University of Eastern Finland
Kuopio, Finland
harri.niska@uef.fi

Pekka Koponen, Member, IEEE
VTT
Espoo, Finland
pekka.koponen@vtt.fi

Antti Mutanen
Tampere University of Technology
Tampere, Finland
antti.mutanen@tut.fi

***Abstract*— Short-term forecasting of electric loads is an essential function required by Smart Grids. Today increasing amount of smart metering data is available enabling the development of enhanced data-driven models for short-term load forecasting. Until now, a plethora of models have been developed ranging from simple linear regression models to more advanced models such as (artificial) neural networks (NNs) and support vector machines (SVMs). Despite the relatively high accuracy obtained, the acceptance of purely data-driven models such as NN models is still remained limited due to their complexity and nontransparent nature. Therefore it is important to develop optimization schemes, which can be used to facilitate the selection of appropriate model structure resulting good forecasting accuracy with low complexity. This study presents an optimization scheme based on multi-objective genetic algorithm (GA) for designing data-driven models for short-term forecasting of electric loads. The optimization scheme is demonstrated for designing the conventional NN/MLP model using real smart metering data and weather forecasts. The optimal NN model structures are identified and analyzed in terms of model complexity and operational forecasting accuracy.**

Keywords—smart metering; data mining; load forecasting; genetic algorithms

I. INTRODUCTION

Short-term forecasting of electric loads is vital for ensuring the functioning of Smart Grids. Network operators require load forecasts for network automation, electricity retailers and aggregators for day-ahead and intraday markets, balance markets, balance management and for scheduling control actions. System operators require load forecasts for scheduling the use of system reserves.

Today, increasing amount of smart metering data is available, covering already up to nearly 100% of all customers in some regions, enabling the development of new, more accurate data-driven modeling schemes for electric load modelling and forecasting [e.g. 2]. Various purely data-based, i.e. data-driven, models have been developed ranging from simple regression models to more advanced models such as neural networks (NNs). Some popular models have been based on linear regression, which decomposes the load into basic and weather, especially outdoor temperature, related components [2], but in recent years also other approaches have been increasingly applied.

Developing appropriate data-driven models is however a complicated process which requires in-depth analysis of various seasonal and exogenous factors, their complex, ill-defined and often time-lagged interconnections, and from this basis finally the selection of optimal model input variables and structure.

Among the data-driven paradigms, NNs for instance have received much attention, and a plethora of papers have reported successful experiments and practical tests with them since the late 1980's [2]. Among the methods multi-layer perceptron (MLP), and in recent years support vector machines (SVMs), have shown to be accurate methods in load forecasting [2, 3]. In general, the advantage of NN models, compared to other statistical methods, is that they can learn complex, non-linear, and a priori unknown relationships from the training data [4]. Nevertheless, the acceptance of NN models has remained limited. Some objections are that NN models are perceived to be highly complex and nontransparent, thus not allowing engineers and system operators to interpret nor understand their behavior properly. Often NN models seem to be over parametrized, meaning that the number of model parameters is so high that model heavily overfits the data and prediction performance degrades [2].

On the basis of aforementioned issues, enhanced methods for selecting appropriate model structure, in particular model input variables, are required in order to improve and justify data-driven model's usability in short-term load forecasting. However, the selection of a feasible model structure is a complex task, due to combinatorial nature of the problem, which makes comprehensive testing of all model structures often difficult or impossible. Thus new enhanced optimization schemes are required to facilitate the model selection. In this context, the methods based on genetic algorithm (GA) are of particular interest, since they have been already shown to have many appealing properties in selection of optimal model architecture in electric load forecasting, e.g. [5].

In this study, we demonstrate a novel multi-objective GA-based scheme for designing simplified data-driven models for short-term forecasting of electric loads. Compared to the previous studies the proposed approach allows an exploration of model set-ups in respect to accuracy and complexity in a reasonable computational time and finally produces a set of model candidates for further analysis and development.

II. Electric Load Forecasting

A. Short-term load forecasting task

The objective was to forecast total hourly power of a large group of small houses and apartments using hourly interval measurements and outdoor temperature forecasts. The forecasting was performed for the next days' hourly powers using the available data in an operational situation at 9 am. The metered data comprised of hourly measured consumption of 3516 customers from the years of 2009 and 2010. The data contained only measurement points where the hourly peak consumption was below 50 kWh.

Measured outdoor temperatures were gathered from the same time period from a measurement point, roughly in the center of the studied power distribution area. An influence of weather forecasting error was simulated by adding the calculated forecasting error to the measured outdoor temperatures. The weather forecasting error was determined based on the measurements and forecasts of the nearest available weather forecast location, located 150 km from the studied network area.

B. Neural network model

The conventional feed-forward multi-layer perceptron (MLP) network with one output was applied [4]. The choice was primarily based on MLP's simplicity and accuracy shown in many application areas, including electric load forecasting [1, 3]. There is a separate comparison of the MLP network with other types of methods such as a partly physically based Kalman-filter, a new load profiling, etc. using the same data [6]. ARIMA methods were omitted from this comparison, because of relatively poor performance so far achieved by us in this case even when a nonlinear input filter was included to deal with nonlinearities. In the comparison the MLP neural network slightly overperformed the other methods in forecasting accuracy. However it should be emphasized the inherent limitations of the standard MLP, and thus the use of more sophisticated modeling techniques is recommended in the further development. In recent years SVM [7], for instance, has emerged as a particularly promising method in modeling and forecasting of electric loads [3].

To briefly describe the well-known basic principles of the MLP network, MLP consists of a network of simple processing elements (neurons) and their connections, arranged in input layer, hidden layers and output layer. Each neuron computes a weighted sum of the inputs, and processes it using a transfer function and distributes the result to the subsequent layer. The output signal y of a single neuron can be expressed as:

$$y = f\left(\sum_{i=i}^{n} w_{ij} x_i + b_j\right) \quad (1)$$

where f denotes the transfer function, j is the index of the neuron, n is the number of neurons in input layer, x_i is the input from ith input neuron, w_{ij} is the weight between ith input neuron and jth hidden neuron and b_j is the bias of the neuron.

The obtained results are highly dependent on the selected network architecture and other parameters, such as the size of the network, transfer functions and training algorithm. Training of the MLP network was performed using the back-propagation (BP) algorithm, also called the generalized (Widrow-Hoff) delta rule and its modifications, such as the Levenberg-Marquardt algorithm, which adjusts iteratively the weights of the network for minimizing the error function, namely the squared errors calculated between actual and desired outputs. At a general level, it is possible to describe the learning using the well-known formula of gradient descent as follows:

$$w(t+1)=w(t)+\alpha(t)g(t) \quad (2)$$

where $\alpha(t)$ is a learning factor, $w(t)$ is a vector of current weights, $g(t)$ is the current gradient for weights and t is a counter for iterations.

The basic structure of the MLP model, to be optimized here, for load forecasting consisted of the hourly power as an output variable, and the timing variables, day length and outdoor temperature as input variables (Table I). The timing variables were divided into sine and cosine components in order to achieve their continuous form. The selected timing variables (day of year, day of week and hour of day) and day length aim at describe major temporal rhythms (hourly, weekly and annual) in the use of appliances and need of lighting, respectively. The outdoor temperature with hourly delays (1−48 hours) aim to describe temperature dependency, fast and slower level dynamics of heating and cooling loads e.g. due to heat storage capacities of buildings, level of isolation, and air conditioning.

TABLE I. MODEL STRUCTURE AND ITS MECHANICAL INTERPRETATION

Model inputs to be optimized	**Sub loads modeled**
Day of year (1-365)	Domestic appliances (seasonal rhythm)
Day of week (1-7)	Domestic appliances (weekly rhythm)
Hour of day (1-24)	Domestic appliances (hourly rhythm)
Day length (hours)	Lighting
Outdoor temperature with time-lags of 1-48 hours (˚C)	Heating and cooling
Model output	
Hourly power (MW)	Hourly power of the target customer group

The MLP network model was trained using Levenberg-Marquardt (LM) algorithm and 3000 training epochs. For controlling over-fitting, the early-stopping was adopted by stopping the training when the error of a test set of training increased for 25 iterations. The selection of an appropriate MLP architecture is a complicated task, for which the specification of exact rules is difficult or even impossible [8]. In this study, the selection of feasible architecture of the network was based largely on experimental testing, which showed that one hidden layer with number of inputs multiplied by two, sigmoid transfer functions for hidden units, and linear transfer function for output is sufficient.

C. Model building and validation

The data from the year 2010 was used as training/identification data and the data from the year 2009 as validation data. The data from the year 2010 was used as training and model optimization set, as it covers larger temperature and power data ranges, and thus allowed us to avoid testing purely data-driven models in data conditions, which had not been experienced through training. Next the extrapolation capabilities will be tested and reported separately by using 2009 for identification and 2010 for validation. Then one of the objectives is to find out to what extent the inclusion of extrapolation affects the selection of the input variables. The temperature range of 2009 was [-24.7 276.] and of 2010 it was [-29.5 32.8].

The performance of the MLP network model was assessed with respect to the standard statistical performance indices (Table II) calculated based on the measured and forecasted total hourly power during the validation year. For more information on measures of forecasting accuracy, the reader is referred to [9]. We also suggest defining and using application specific performance measures, when the specific applications and their priorities are known.

TABLE II. PERFORMANCE MEASURES FOR MODEL VALIDITY

Performance measure	**Equation**
Root Mean Square Error (RMSE)	root(mean($e_t^{\,2}$))
Mean Absolute Error (MAE)	mean ($\lvert e_t \rvert$)
Mean absolute percentage error (MAPE) that is the sum of absolute errors divided by the sum of observed values	mean($\lvert p_t \rvert$), where p_t= 100 e_t / y_t ,where y_t is the observation at time t
Sum of Squared Error of prediction (SSE)	sum($e_t^{\,2}$))
Here e_t is the forecasting error at time t. For more information on measures of forecast accuracy, the reader is referred to [9].	

III. MODEL SELECTION USING GENETIC ALGORITHM

A. *Proble formulation and objectives*

The selection of an optimal model input subset for the short-term load forecasting task was formulated as a multi-objective optimization problem with two optimization criteria to be minimized:

(i) the number of model input variables

(ii) the forecasting error of model

As opposed to the single-objective optimization, a set of trade-off solutions, called Pareto-optimal/non-dominated solutions, are achieved, which cannot be improved in any of the objectives without sacrificing at least one of the other objectives. The proposed multi-objective optimization problem was then solved using genetic algorithm (GA) [10, 11] by encoding the model inputs (see Table I) using the binary coding presentation.

To evaluate the fitness of candidate model structures, the accuracy of the model for a specific input subset was measured using the standard, scale-independent Mean Absolute Percentage Error (MAPE) index (see Table II). However, the accuracy index should be selected based on the requirements of the forecasting application. In load forecasting MAPE prefers high accuracy during low load while in power systems accuracy during high/peak loads is usually much more important. Therefore, the optimization of model structure using other performance measures is of interest in further studies.

B. *Sensitivity analysis*

The problem of the proposed optimization scheme is however with the computational burden of the fitness assessment, which makes comprehensive testing of all model structures (i.e. input subsets) difficult, impossible or at least impractical. Therefore, following [12], the sensitivity analysis of the model was used instead of the actual, time-consuming model fitting/training to indicate the predictive power of input subsets. In the scheme, the complete model is initially trained using all the potential input variables of training data, which are then simulated using a set of the training data where the absence of inputs are replaced by their respective means, computed on the training set. The proposed approach can be used to estimate the predictive power of a subset of input variables, instead of one input variable at a time. Consequently, the accuracy of a model candidate is measured as by means of a sensitivity *s* of an input subset, i.e. the error difference (MAPE):

$$s=MAPE_{train}-MAPE_{sim} \tag{3}$$

where $MAPE_{train}$ is an error of the model trained using the all candidate inputs and $MAPE_{sim}$ is an error achieved by simulating the model for the input vector with absence inputs.

C. *GA implementation and search parameters*

The multi-objective GA within the unrestricted migration/island model [10, 11] provided by the Matlab GEATbx toolbox was adopted for evolving the NN/MLP model input variables. In this scheme, the 5 subpopulations, the base population of 100 individuals and 150 generations were used, by exchanging 10% of individuals between subpopulations within 20 generations. Different subpopulation specific mutation rates were employed in order to maintain rough and fine search capabilities. The selection of operators was based largely on the investigations made by [13], which show the advantages of the random sampling of tournament selection and the subset size-oriented common features (SSCOF) recombination in multi-objective input selection. SSCOF is capable of constructing useful building blocks and maintaining the distribution across the range of Pareto front.

IV. OPTIMIZATION RESULTS

A. *Optimal NN model structure*

The computed Pareto-optimal fronts, i.e. model structures, regarding to the objectives: the model accuracy (measured using the sensitivity) and the model complexity (the number of inputs) obtained during the multi-objective GA run are depicted in Fig. 1. In general, sufficient convergence is indicated for 150 generations adopted. The optimization scheme seems to find feasible solution space during the first 30 generations. After that, the fine search capabilities of the parallel GA ensure the identification of the final Pareto-front to be further analyzed and validated.

There is a risk that the obtained Pareto-optimal model structures overfit to the identification data, and thus do not provide a good generalization on an external data. Therefore the obtained Pareto-optimal model structures were validated using external test data excluded from the identification set. The optimality of the model structures regarding to the sensitivity and the validated performance is depicted in Fig. 2. In addition, the numerical results as well as the selected inputs are given in Table III.

In general, the results show that 7−9 input variables are required to achieve maximal external forecasting performance.

Increasing the number of input variables over 10 does not enhance forecasting accuracy, but rather decreases the external prediction power. Among others, the resulting optimal input variables are all the timing variables, the length of day and outdoor temperature with specific hourly time-delays (see Table III).

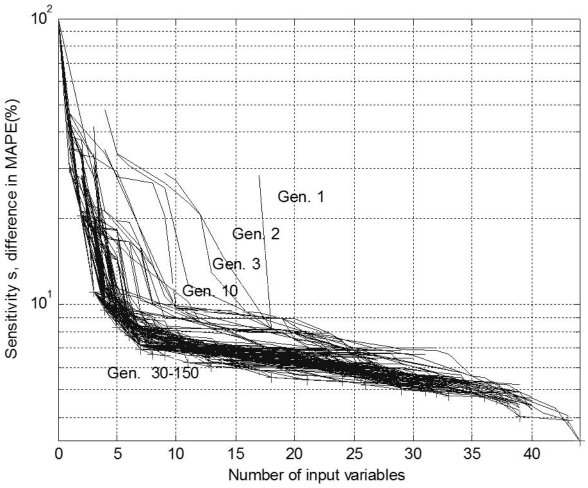

Fig. 1. The convergence of multi-objective GA in terms of resulting Pareto-fronts during 150 generations.

The main uncertainty is related to the selection of appropriate time-delays for hourly outdoor temperatures, which needs further consideration. In general, the results (Table III) show that multiple delayed temperature values are required in the modeling to achieve sufficient forecasting accuracy. The time-lag of 9 hours tends to describe the significant part of heating and cooling load dynamics. However, according to the results (Table III) also longer temperature time lags (such as 34/41 hours) are needed to describe slower dynamics.

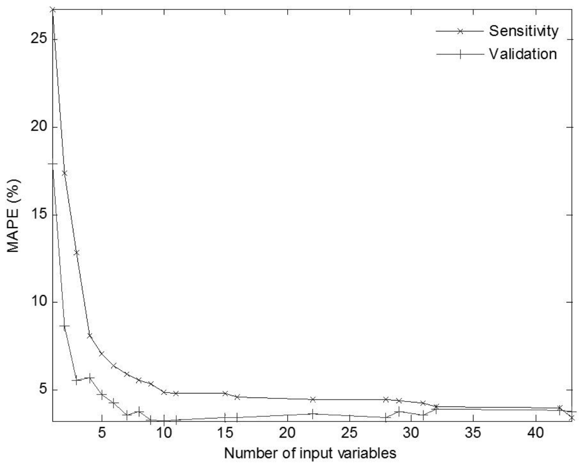

Fig. 2. Optimality of the NN model in respect to number of input variables and forecasting error (MAPE), measured using the sensitivity (marked with "x") and the external validation (marked with "+").

TABLE III. PARETO-OPTIMAL NN MODEL STRUCTURES, SENSITIVITY AND EXTERNAL PERFORMANCE

Number of inputs	**Timing variables**			**Day length**	**Outdoor temperature at T+0hour**	**Number of delayed T (in brackets delay lengths (hours))**	**Sensitivity (ΔMAPE%)**	**External performance (MAPE%)**
	Day of year	**Day of week**	**Hour of day**					
1					x		26.8	17.9
2			x		x		17.4	8.6
3	x		x		x		12.8	5.6
4	x		x	x	x		8.1	5.6
5	x	x	x	x	x		7.0	4.7
6	x	x	x	x	x	1 (9)	6.4	4.2
7	x	x	x	x	x	2 (9, 34)	5.9	3.5
8	x	x	x	x	x	3 (5, 8, 11)	5.5	3.7
9*	x	x	x	x	x	4 (8, 9, 13, 41)	5.3	3.3
10	x	x	x	x	x	5 (5, 8, 9, 13, 45)	4.8	3.3
11	x	x	x	x	x	6	4.8	3.3
15	x	x	x	x	x	10	4.8	3.4
16	x	x	x	x	x	11	4.6	3.5
22	x	x	x	x	x	17	4.4	3.7
28	x	x	x	x	x	23	4.4	3.6
29	x	x	x	x	x	24	4.4	3.8
31	x	x	x	x	x	26	4.2	3.5
32	x	x	x	x	x	27	4.0	4.0
42	x	x	x	x	x	37	3.9	4.0
43	x	x	x	x	x	38	3.4	3.8

*Selected model for the final evaluation and comparison

B. *Forecastig accuracies*

The resulting external power forecasts of the optimized NN model (the model with 9 input variables) for the validation year are illustrated in terms of measured versus forecasted hourly powers (Fig. 3) and measured and forecasted hourly power and error residual time-series (Fig. 4).

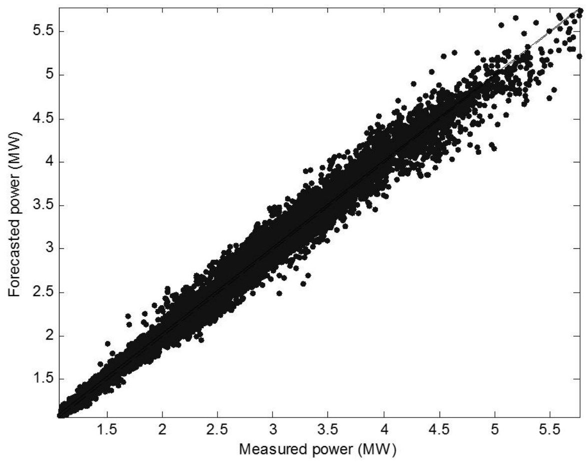

Fig. 3. Plots of measured versus forecasted hourly powers (MW) by the optimized NN model.

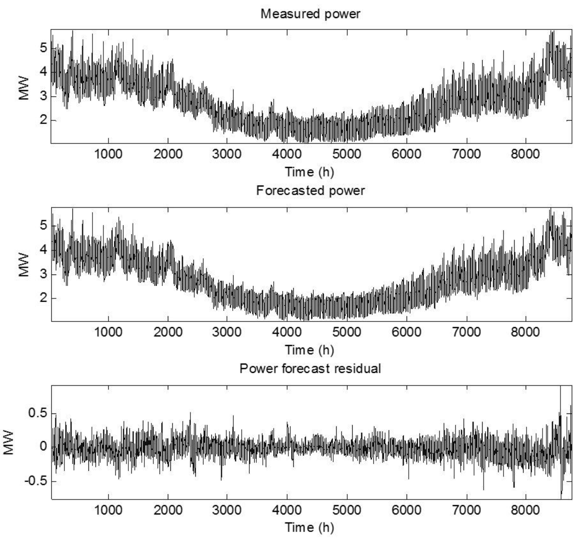

Fig. 4. Measured and forecasted hourly powers and respective error residuals during the validation year.

In addition, the statistical measures were calculated for the performance of the NN model (Table IV). According to Table IV high external forecasting performance was achieved (MAPE 3.4%) without significant bias. Examination of the resulting power forecast residuals (Fig. 4) however shows significant peaks in the end of the validation year, indicating that holidays (Christmas) may not be properly taken into account by the proposed NN model.

To analyze the influence of weather forecast error on the performance and to get an understanding about the true operational performance, the validation was also done with hourly measurements of outdoor temperatures instead of the simulated temperature forecasts. According to this test (Table IV), MAPE seems to decrease roughly 0.3% from 3.4% to 3.1% when using the actual temperature measurements instead of temperature forecasts based on the operational weather forecast model, which is not very significant in an operational sense. However, the analysis of an impact of weather forecasting error should be carried out more extensively using real site-specific weather forecast data in order to get a better understanding of its influence and possible risks. The results of this matter are pending and will be published separately.

TABLE IV. FORECASTING PERFORMANCE ACHIEVED USING THE OPTIMIZED ANN MODEL

Model	**Performance measure**			
	std (MWh)	**MAE (MWh)**	**MAPE (%)**	**SSE (MWh^2)**
Daily energy	1.92	1.42	2.05	1437.36
Hourly energy	128.37 (115.18)	95.63 (85.01)	3.42 (3.06)	151.48 (117.39)

*in brackets the performance achieved using the measured temperature

V. CONCLUSION

As a conclusion, the multi-objective optimization scheme based on GA was demonstrated for designing data-driven models for short-term forecasting of electric loads. In general, the selection of model input variables in terms of model complexity, interpretation and accuracy, is a tedious task with purely data-based models [1]. The proposed optimization scheme mitigates this problem by identifying and removing redundant parts of model structures. The benefit of the proposed scheme is particularly on its capability for global exploration of simplified model structures in respect to multiple criteria, simultaneously.

As a recommendation for future work, it is necessary to test the scheme for other data-based modeling techniques such as SVR as well. In addition, more extensive validation using more extensive smart metering data is required in order to achieve wider understanding about the performance and limitations of the approach.

VI. ACKNOWLEDGMENT

This research was supported by the Finnish Smart Grids and Energy Markets (SGEM) research program of the Cluster for Energy and Environment (CLEEN) and the Response project of the Academy of Finland. The authors would like to thank the distribution operator KSS (Koillis-Satakunnan Sähkö Oy) and its customers for providing the measurement data.

REFERENCES

[1] Räsänen, T., Voukantsis, D., Niska, H., Karatzas, K., and Kolehmainen, M. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. *Applied Energy*, vol. 87, pp. 3538-3545, 2010.

[2] Hippert, H. S., Pedreira, C. E., and Souza, R. C. Neural Networks for Short-Term Load Forecasting: A Review and Evaluation. *IEEE Transactions on Power Systems*, vol. 16, no. 1, 44-55. 2001.

[3] Hahn, H., Meyer-Nieberg, S. and Pickl, S. Electric load forecasting methods: tools for decision making. *European Journal of Operational Research*, vol. 199, pp. 902–907, 2009.

[4] Haykin, S. *Neural Networks: A Comprehensive Foundation*. 2nd Edition, Prentice-Hall, Upper Saddle River, NJ, 1999.

[5] Tian, L. and Noore, A. Short-Term Load Forecasting Using Optimized Neural Network with Genetic Algorithm. In *Proc. of 8th International Conference on Probabilistic Methods Applied to Power Systems,* pp. 135-139, 2004.

[6] Koponen, P., Mutanen, A., and Niska, H. Assesment of Some Methods for Short-Term Load Forecasting. In *Proc. of the 5th IEEE PES Innovative Smart Grid Technologies (ISGT) European 2014 Conference*, 12-15 October, 2014, Istanbul, Turkey, IEEE, Piscatawa, NJ. http://dx.doi.org/10.1109/ISGTEurope.2014.702890.

[7] Vapnik, V. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.

[8] Niska, H., Hiltunen, T., Karppinen, A., Ruuskanen, J., and Kolehmainen, M. Evolving the neural network model for forecasting air pollution time series. *Engineering Applications of Artificial Intelligence*, vol. 17, pp. 159-167, 2004.

[9] Hyndman, R. J. and A.B. Koehler, A. B. Another look at measures of forecast accuracy. Monash University, 2005. http://www.robjhyndman.com/ papers/mase.pdf

[10] Srinivas, N. and Deb, K. Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, Vol. 2, pp. 221-248, 1994.

[11] Fonseca, C. M. and Fleming, P. J. Genetic algorithms for multi-objective optimization: formulation, discussion and generalization. In *Proc. of the 5th International Conference on Genetic Algorithms*, San Meteo, pp. 416-423, 1993.

[12] Niska, H., Rantamäki, M., Hiltunen, T., Karppinen, A., Kukkonen, J., Ruuskanen, J., and Kolehmainen, M. Evaluation of an Integrated Modelling System Containing a Multi-Layer Perectpron Model and the Numerical Weather Prediction Model HIRLAM for the Forecasting of Urban Airborne Pollutant Concentrations. *Atmospheric Environment*, vol. 39, pp. 6524-6536, 2005.

[13] Emmanouilidis, C., Hunter, A., and MacIntyre, J. A multi-objective evolutionary setting for feature selection and a commonality-based crossover operator. In *Proc. of CEC00, Congress on Evolutionary Computation*, IEEE, Piscataway, NJ., USA, pp 309-316, 2000.