Predicting Air Temperature and Relative Humidity Using a Statistical Inductive Learning Simulator

Mansour Nasser Jadid

Department of Building Science &Technology,
King Faisal University
P. O. Box 30973 Al-Khobar 31952, Kingdom of Saudi Arabia

Abstract

This paper present a Statistical Inductive Learning Simulator (SILS) to forecast the maximum daily temperature (maxdt) and the maximum daily relative humidity (maxrh) for the Saudi electrical company in Eastern Province. Predicting and visualizing engineering parameters is an essential issue for the electrical company to search for modern ways to cut losses due to the extreme consumption of electrical energy during summer season.

The weather data that consist of 36 daily weather records (parameters) were provided by Meteorological Eastern Province of Saudi Arabia (METEP). SILS is a learning Abduction Induction Machine (AIM) which automatically models the data by training and evaluation. The result is an automatic-function (input-output relationship) based on a statistical model. The study has stressed the use of different model settings such as Complexity Penalty Multiplier (CPM), number of layers, and size of layers which is a part of the program. The data has been trained for different number of parameters and manipulated by either destruction or construction models until optimum model was produced.

The predicted daily weather data of maximum daily temperature 'maxdt' and the maximum ralative humadity (maxrh) was provided to synthesizing model for four years (1990,1991,1993,1994) as training. The desired data for the 'maxdt' and the (maxrh) of any day in year 1992 used for predictions. Error percentage was calculated based on the estimated and actual data.

Key words: Statistical Inductive Learning, Visualization, Complexity Penalty Multiplier, Prospector, Maximum Daily Temperature, Maximum Ralative Humadity, Synthesis, Traning, Evaluation.

Introduction

Application of artificial intelligence has gained worldwide acceptance. The artificial intelligence falls into Knowledge-Based (Experts) Systems or Machine Learning Techniques or combination of both. Treamodus research works have been carried out in the engineering applications using artificial neural networks (Jadid 2000, Jadid and Fairbain, 1996, 1998, SubbaNarasimha, et al. 2000)

SILS (ModelQuest Prospector, 1997) is a machine learning tool that automatically abducts the network solutions to complex designs, prediction, and controlling. In addition to synthesizing statistical networks, it encodes them into computer subroutines that can be linked with the main C program. Statistical networks can accurately predict complex function and substitute relationships that cannot be obtained by conventional statistical method. SILS provides a complete environment that analyze, synthesize, and encode statistical network.

Local electrical companies always look for modern ways to cut losses due to the extreme consumption of the electrical energy. In summer season between May and September the Saudi electrical company in the eastern province recorded many losses and occasionally shutdown of power plant due to high electricity consumption. The company major facilities (substations and transformers and power plants) required to be maintained under certain temperature. Researchers, electrical engineers designers and planners need simulated climatic conditions, especially, air temperature and relative humidity during the initial design. Therefore, there is always a need for using artificial intelligence such as learning-machine or expert's system application, to assist in predicting temperature and humidity for the operation.

A historical data was used to build a prediction model. The main objective was to produce reliable and cheap model for the electrical company for both maximum daily temperature and maximum relative humidity for any day of year 1992 by using climatic from years data between 1990 and 1994. The daily weather data was arranged for synthesizing process into three sets of parameters (9,18 and 36) for different number of years. The predicted data has been evaluated, tested and errors had been recorded in percentages.

Mathematical Modelling

SILS is a statistical neural network tool, which trains a set of examples, and predicts mathematical relationship in a set of data. It is an excellent tool in modeling numeric information; for example it can perform the following modeling:

1. Automatically discovers network solution of a complex problem from a set of examples.
2. Maps a mathematical model of the relationship in the data.
3. Automatically determines the best structure model, element types, coefficients and correlations based on the selection of the most accurate model without over-fitting the data.

The module produced by the synthesis process is powerful and compact transformation implemented as a layered network of feed-forward functional elements. The output of any given element, together with the original input variables, can feed into subsequent layers. Models are trained from layer to layer until the model ceases to improve. The *P*redicted *S*quare *E*rror (*PSE*) (Barron, 1994) is the modeling criterion. It is a heuristic measure of the expected network squared error for independent data set. The *PSE* is derived from:

$$PSE = FSE + KP \qquad (1)$$

Where:

PSE = *P*redicted *S*quare *E*rror

FSE = *F*itting *S*quare *E*rror of the model for the training data.

KP = complexity penalty

Where;

$$KP = CPM * (2K / N) Sp^2 \qquad (2)$$

Where;

K = the total number of coefficients.

N = the number of training data.

Sp² = a-priori estimate of the true unknown model error variance.

The constants *K*, *N*, and *Sp²* are determined by the set of examples which used to train the model. As *Sp²* decreases or *N* increases, it has the ability of fitting the data with more confidence and allow for more *C*omplexity *P*enalty *M*ultiplier (*CPM*). The *CPM* value (ModelQuest Prospector™, 1997) can be adjusted. The default value for the *CPM* allows using its best estimate for *KP* value. The *CPM* is a variable that can be adjusted within the tool. As the value for the *CPM* increases the impact of the complexity penalty term increases.

The default sitting values of the *C*omplexity *P*enalty *M*ultiplier (*CPM*), the number of layers and the size of layer are 1, 4 and 15 respectively. Several models were produced which are based on mathematical equations (Montgomery, Drake, 1990):

Single element:

$$w_0 + (w_1 x_1) + (w_2 x_1^2) + (w_3 x_1^3) \tag{3}$$

Doublet elements:

$$\begin{aligned} &w_0 + (w_1 x_1) + (w_2 x_2) + (w_3 x_1^2) + (w_4 x_2^2) + (w_5 x_1 x_2) + (w_6 x_1^3) + \\ &(w_7 x_2^3) + (w_8 x_2 x_1^2) + (w_9 x_1 x_2^2) \end{aligned} \tag{4}$$

Triplet elements:

$$\begin{aligned} &w_0 + (w_1 x_1) + (w_2 * x_2) + (w_3 * x_3) + (w_4 * x_1^2) + (w_5 * x_2^2) + (w_6 * x_3^2) + \\ &(w_7 * x_1 * x_2) + (w_8 * x_1 * x_3) + (w_9 + x_2 * x_3) + (w_{10} * x_1 * x_2 * x_3) + \\ &(w_{11} * x_1^3) + (w_{12} * x_2^3) + (w_{13} * x_3^3) + (w_{14} * x_2 * x_1^2) + (w_{15} * x_1 * x_2^2) + \\ &(w_{16} * x_1 x_3^2) + (w_{17} * x_3 * x_1^2) + (w_{18} * x_3 * x_2^2) + (w_{19} * x_2 * x_3^2) \end{aligned} \tag{5}$$

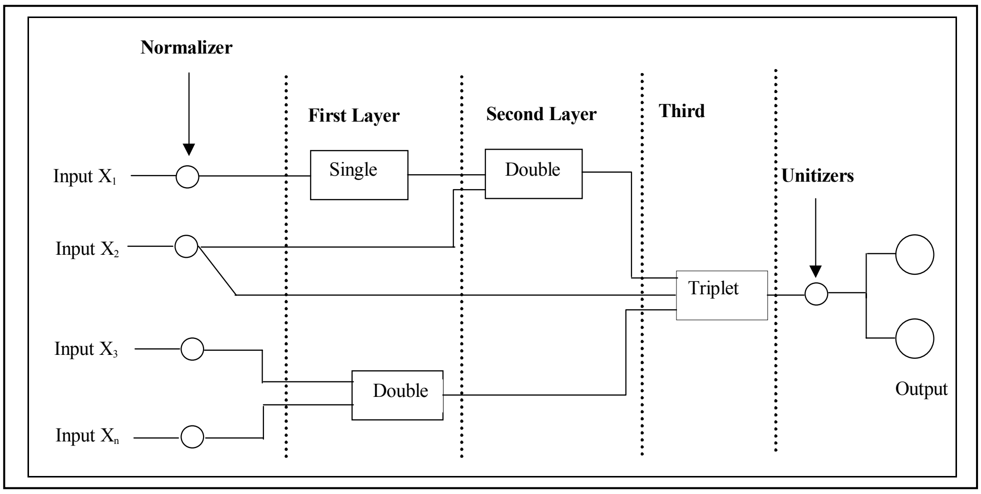
Where;

w_n = coefficients determined by SILS.

x_n = input variables.

Single, doublets and triplets are elements with names based on the number of input variables. These elements are of third degree polynomial equations and doublet and triplets have cross-terms that allow interaction between the nodal input variables. A typical network structure used is shown in Figure (1).

FIGURE (1) Three layers statistical inductive learning network



Simulator Application and Operation

There are many areas that can be used to develop practical application models to help decision-maker, instructor, planner, marketer, predictor and developer in solving certain problems or helping in making decision.

There are only four major steps needed for application:

1. **Import data**: data should be imported in the form of ASCII or rich-text files.
2. **Train data**: manipulation and training of the data before synthesizing the model. The variables could be specified as input-output; and the training could be fine-tuned by adjusting the setting and generation of the data model parameter with different values.
3. **Evaluation and performance**: the unseen tested data could be evaluated. Histogram and specification of performance of the model could be obtained as graph or text file.
4. **Implementing the model**: once the data used to be trained and model obtained, the results can be obtained by the query function as interactive tools or by converting the model into 'C' subroutine.

Process for Data Preparation

The data for 36 daily weather elements were exported to the program and synthesized by the network. The trained data were divided into 3 sets of parameters as follow:

Model (A) for Nine (9) Set Parameters of Input Data

Model (B) for Eighteen (18) Set Parameters of Input Data

Model (C) for Thirty Six (36) Set Parameters of Input Data

Different models have been generated for each set of parameters. The best model is the one that scores more in the synthesizing criteria. Outputs are maximum daily temperature (maxdt) and maximum relative humidity (maxrh) while the input are the rest of weather data for each parameters set. Manipulating model, synthesizing setting and data generation have been manipulated by selecting best models within the simulator. Once the model is set with different value of CPM, number of layers and size of layers network then it synthesizes automatically. Predicted models were developed after testing and evaluations.

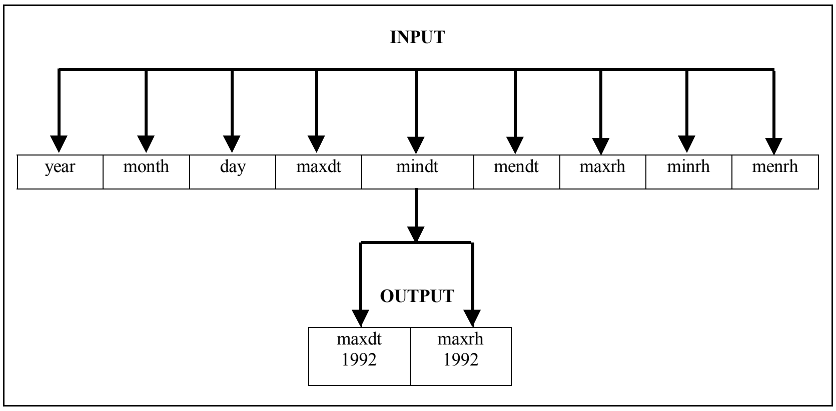
Modeling Maximum Daily Temperature and Relative Humidity

The data were divided into three different sets of parameters, (9, 18, and 36). The models produced from each set of parameters data were manipulating by different synthesizing setting (CPM value, number of layers and size of layers) and trained by percentage of the total data 75% and 95%. The details and explanation of the all sets of parameters and their statistical criteria are explained as:

Model (A) for Nine (9) Set Parameters of Input Data

In this model, the 1826 daily weather observations representing 9 parameters (year, month, day maxdt, mindt, mendt, maxrh, minrh, menrh), were used as input to the network synthesis procedure. Outputs are the maximum daily temperature (maxdt) and maximum relative humidity (maxrh). Figure 2 shows input-output relationships for model (A):

FIGUR 2. Input-output relationships for model (A)



Each model of the 9 parameters set was created with different synthesizing setting and generated with 75% and 95% of the total training data. Table 1 and 2 show the performance of the six plus optimum predicting models and their statistical criteria for both maximum daily temperature and maximum daily humidity.

TABLE 1. Performance for Models (A): Six plus Optimum for Maximum Daily Temperature

Percentage of Total Data	CPM	Number of Parameters	Model Synthesize Settings				Average Absolute Error	Absolute Error Standard Deviation	Average Squared Error	Squared Error Standard Deviation	Maximum Absolute Error
			CPM Value	Number of Layers	Size of Layers	Time					
75%	Default	9	1	4	15	5 sec	0.054815	0.11414	0.016006	0.2898	2.458
	Minimum	9	0.01	1	2	1 min	0.054834	0.11414	0.016006	0.28287	2.458
	Maximum	9	10	9	100	20 sec	0.054815	0.11417	0.016012	0.28287	2.458
95%	Default	9	1	4	15	5 sec	0.049217	0.0088731	0.0025001	0.00088725	0.061022
	Minimum	9	0.01	1	2	4 sec	0.054834	0.11414	0.016006	0.28287	2.458
	Maximum	9	10	9	100	2 min	0.049217	0.0088731	0.0025001	0.00088725	0.061022
	Optimum	9	0.01	4	10	5 sec	0.059919	0.0018205	0.00034979	0.00021528	0.060416

TABLE 2. Performance for Models (A): Six plus Optimum for Maximum Relative Humidity

Percentage of Total Data	CPM	Number of ParameterS	Model Synthesize Settings				Average Absolute Error	Absolute Error Standard Deviation	Average Squared Error	Squared Error Standard Deviation	Maximum Absolute Error
			CPM Value	Number of Layers	Size of Layers	Time					
75%	Default	9	1	4	15	5 sec	0.49936	0.073554	0.255476	0.11394	1.5654
	Minimum	9	0.01	1	2	5 sec	0.50008	0.075505	0.25577	0.11384	1.5497
	Maximum	9	10	9	100	10 sec	0.49936	0.073554	0.25476	0.11394	1.5654
95%	Default	9	1	4	15	5 sec	0.49785	0.023225	0.24839	0.023143	0.54261
	Minimum	9	0.01	1	2	4 sec	0.50008	0.033247	0.25577	0.032619	0.58637
	Maximum	9	10	9	100	2 min	0.49785	0.023225	0.25476	0.02314	0.54261
	Optimum	9	0.01	4	10	5 sec	0.53863	0.019938	0.29033	0.021479	0.55273

It was observed from Table (1) for maximum daily temperature with 75% of the total traning data that the synthesizing setting was not influenced the statistical criteria. It's seen that the maximum absolute error was 2.458 for all model synthesized with different CPM values, number of layers and size of layers. However, there was good improvement when the model trained with 95% of the total training data except in the minimum model setting where a value of 2.458 was observed. The performance for the maximum relative humidy as shown in Table 2, was much better and values of the maximum absolute error were reduced. However, increasing the total percenatge of training data values of maximum error were reduced; and therefore more percentage of data prove better results. The 75% of the total trained data consist of 1370 while the 95% consist of 1437.

Figure 3 and 4 show the performance of prediction for the default models for the first 31 days of January 1992.

FIGURE 3. Default model (A) performance synthesized
(CPM=1, Number of Layers = 4, Number of Layer Sizes = 15)
generated by 75% of training total data.

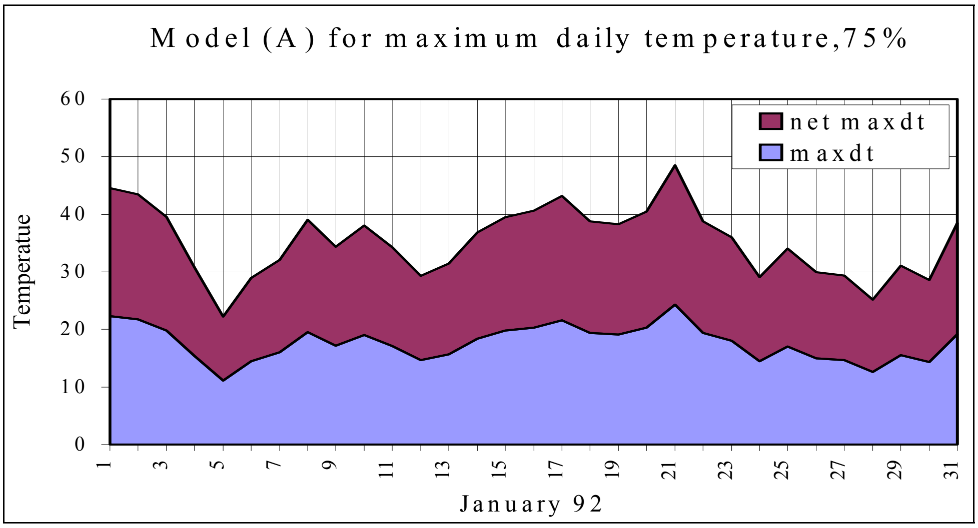
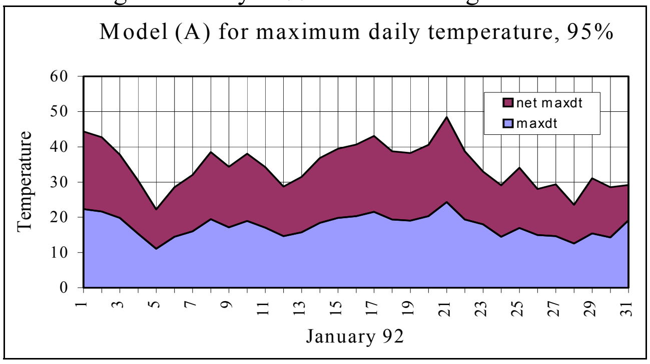


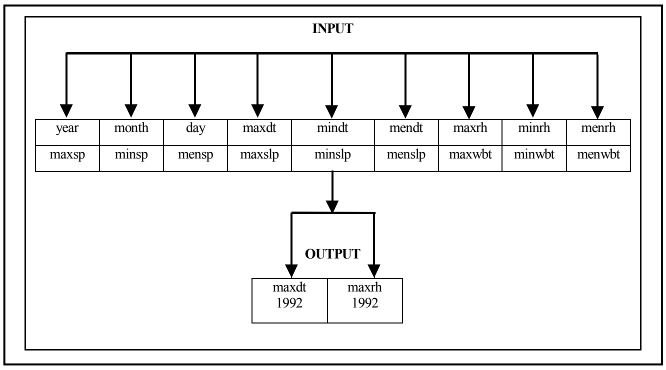
FIGURE 4. Default model (A) performance synthesized (CPM = 1 Number of Layers = 4, Number of Layer Sizes = 15) generated by 95% of total training data.



Model (B) for Eighteen (18) Set Parameters of Input Data

The 1826 daily weather observations represent 18 climatic parameters (year, month, day, maxdt, mindt, mendt, maxrh, minrh, menrh,.... etc), were used as inputs. Outputs are the maximum daily temperature (maxdt) and maximum relative humidity (maxrh). Figure 5 shows input-output relationship for model (B).

Figure 5. Input-output relationships for model (B)



The 18 parameters data model was trained with different setting for 75% and 95% of the total data. Table 3 and 4 show the performance of the six predicting models and their statistical criteria of model (B) for both maximum daily temperature (maxdt) and maximum daily humidity (maxrh).

TABLE 3. Performance for Six Models (B) for Maximum Daily Temperature

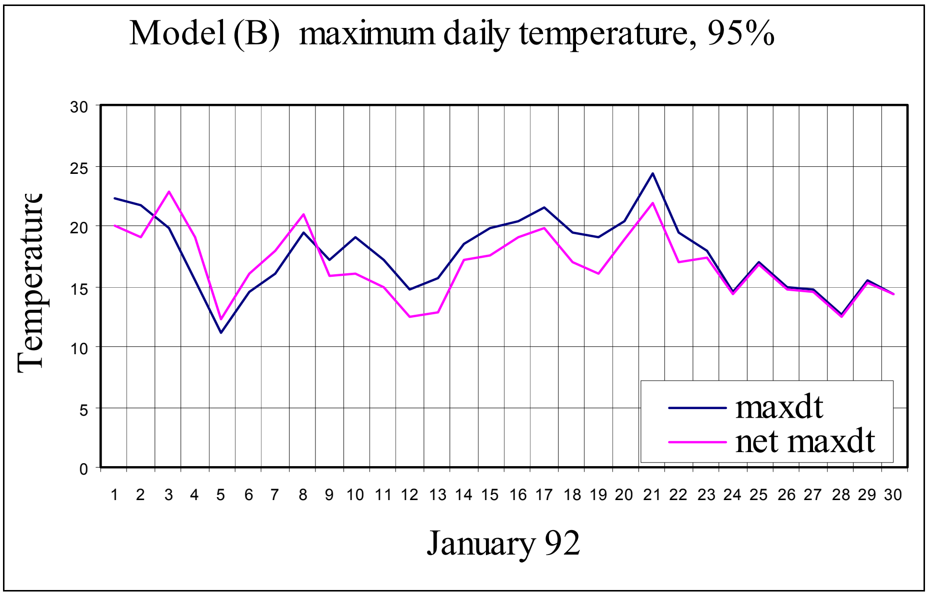
Percentage of Total Data	CPM	Number of Parameters	Model Synthesize Settings				Average Absolute Error	Absolute Error Standard Deviation	Average Squared Error	Squared Error Standard Deviation	Maximum Absolute Error
			CPM Value	Number of Layers	Size of Layers	Time					
75%	Default	18	1	4	15	20 sec	0.054815	0.11417	0.016012	0.2898	2.4585
	Minimum	18	0.01	1	2	20 sec	0.085091	0.17799	0.038851	0.31069	2.4683
	Maximum	18	10	9	100	20 sec	0.083733	0.17799	0.038623	0.31125	2.472
95%	Default	18	1	4	15	10 sec	0.049193	0.008967	0.0024995	0.00089668	0.060827
	Minimum	18	0.01	1	2	6 sec	0.049984	0.0086096	0.0025715	0.00086304	2.060827
	Maximum	18	10	9	100	2 min	0.071132	0.12352	0.020149	0.10097	0.06112

TABLE 4. Performance for Six Models (B) for Maximum Relative Humidity

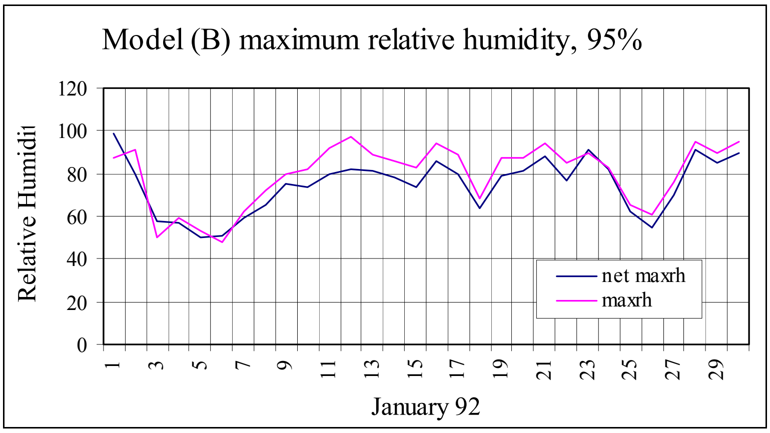
Percentage of Total Data	CPM	Number of Parameters	Model Synthesize Settings				Average Absolute Error	Absolute Error Standard Deviation	Average Squared Error	Squared Error Standard Deviation	Maximum Absolute Error
			CPM Value	Number of Layers	Size of Layers	Time					
75%	Default	18	1	4	15	20 sec	0.49936	0.073554	0.25476	0.11394	1.5654
	Minimum	18	0.01	1	2	20 sec	0.49764	0.092108	0.25611	0.1606	1.847
	Maximum	18	10	9	100	20 sec	0.49936	0.73554	0.25476	0.11349	1.5654
95%	Default	18	1	4	15	10 sec	0.49785	0.023225	0.24839	0.023143	0.54261
	Minimum	18	0.01	1	2	6 sec	0.49488	0.033247	0.246	0.032619	0.58673
	Maximum	18	10	9	100	2 min	0.49785	0.023225	0.24839	0.023143	0.54261

From Table 3, it was observed that the maximum error for the maximum daily temperature increased from 0.0604 when the data model trained was with the 95% of the data for model A to 0.06112 when data model was trained with 95% of the data for model (B). The maximum absolute errors for both the maximum daily temperature and the maximum relative humidity were significantly reduced using 95% of the total traning data. Figure 6 and 7 show performance comparison for the best two predicting (B) models of default CPM values for maximum daily temperature (maxdt) and maximum daily relative humidity (maxrh) respectively.

FIGURE 6. Default model (B) performance for maxdt (CPM = 1, Number of Layers = 4, Number of Layer Sizes =15) generated by 95% of total training data.



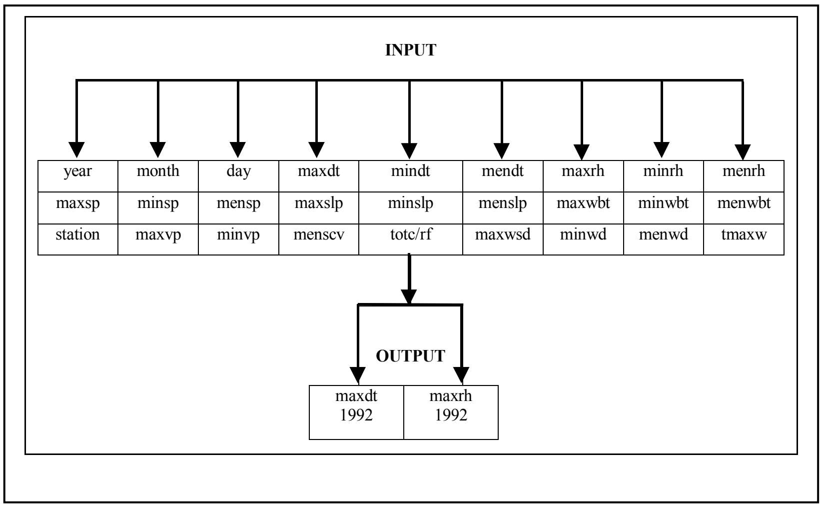
FIGUR 7. Default model (B) performance maxrh (CPM =1, Number of Layers = 4, Number of Layer Sizes =15) generated by 95% of total training data



Model (C) for Thirty Six (36) Set Parameters of Input Data

The 1826 daily weather data observations for all 36 set of parameters (year, month, day, maxdt, mindt, medt, maxrh, minrh, menrh, …. etc.) were used as inputs. Outputs are the maximum daily temperature (maxdt) and maximum relative humidity (maxrh). Figure 8 shows the representation of a input-output relationship.

FIGURE 8. Input-output relationships for model (C)



The 36 parameters data models were trained with the same synthesizing setting for the 36 parameters and data trained with 75% and 95% of total data. Table 5 and 6 show the performance comparison of the six predicting models and their statistical criteria for both maximum daily temperature (maxdt) maximum daily humidity (maxrh), respectively.

From Table 5, for the maxdt, the maximum error reached to 40408% for shorter time duration (8 sec or 3 min), while it reached 1150% for longer time (3 hours) for 95% of the total data. From Table 6, for maxrh, it was observed that the maximum error reached up to 1000% or different time and percentage of the total data.

From the Tables 5 and 6 it was observed that the statistical criteria deteriorated and the error was extremely large (reached up to 1000% for the absolute maximum error). The time of synthesizing of model took 4 hours. There was slight improvement in the absolute average error where it's decrease from 61.018 for the 75% total data to 31.987 when model trained with 95%.

Evaluation

The overall evaluations of the three models are shown in Table 7. It was found that the 9 parameters predicting model score the best statistical criteria. A linear relationship for input-output was produced in the 9 parameter prediction model (A). A nonlinearly parameters was introduced for models B and C. SILS seek to set functioning relationship with the variables been added. However, when the SILS tried to fit curve within the given inputs-output of a doublet and triplet relationships that were introduced in the network synthesizing, confusion took place as result the input-output functioning relation was deteriorating in model (B), and (C). Figure 9 shows the elements relationship of each prediction model.

TABLE 5. Performance for Six Models (C) for Maximum Daily Temperature

Percentage of Total Data	CPM	Number of Parameters	Model Synthesize Settings				Average Absolute Error	Absolute Error Standard Deviation	Average Squared Error	Squared Error Standard Deviation	Maximum Absolute Error
			CPM Value	Number of Layers	Size of Layers	Time					
75%	Default	36	1	4	15	8 sec	1387.8	7150.7	5.25E+07	2.86E+08	40408
	Minimum	36	0.01	1	2	3 min	339.37	3233.3	1.05E+07	1.29E+08	40408
	Maximum	36	10	9	100	20 min	77.707	145.75	27233	1.11E+05	1174.1
95%	Default	36	1	4	15	8 min	1687.8	7150.7	5.25E+07	2.86+08	40408
	Minimum	36	0.01	1	2	3 min	107.77	227.41	62761	2.61E+05	1503.8
	Maximum	36	10	9	100	3 hrs	70.707	138.75	27200	1.11E+05	1150

TABLE 6. Performance for Six Models (C) for Maximum Relative Humidity

Percentage of Total Data	CPM	Number of Parameters	Model Synthesize Settings				Average Absolute Error	Absolute Error Standard Deviation	Average Squared Error	Squared Error Standard Deviation	Maximum Absolute Error
			CPM Value	Number of Layers	Size of Layers	Time					
75%	Default	36	1	4	15	8 sec	28.462	102.06	1.12E+04	9.59E+04	1000
	Minimum	36	0.01	1	2	3 min	28.462	102.06	1.12E+04	9.59E+04	1000
	Maximum	36	10	9	100	20 min	61.018	130.11	20614	9.68E+04	1000
95%	Default	36	1	4	15	8 min	32.987	106.54	1.23E+04	1.05E+05	1000
	Minimum	36	0.01	1	2	3 min	28.462	102.06	11203	1.81E+03	1000
	Maximum	36	10	9	100	3 hrs	32.987	106.54	12315	1.05E+05	1000

TABLE 7. Best Performance from Model (A) for the Maximum Daily Temperature and Maximum Relative Humidity.

Percentage of Total Data	CPM	Number of Parameters	Model Synthesize Settings				Average Absolute Error	Absolute Error Standard Deviation	Average Squared Error	Squared Error Standard Deviation	Maximum Absolute Error
			CPM Value	Number of Layers	Size of Layers	Time					
maxdt	**Optimum**	9	0.01	4	110	5 sec	0.059919	0.0018205	0.00034979	0.00021528	0.060416
maxrh	**Default**	9	0.01	1	2	5 sec	0.49785	0.023225	0.24839	0.023143	0.54261

It was observed that the 9 parameters set produced the optimum model (model A). Manipulating the synthesis setting and data training percentage (75% and 95%) produced the improvement of the optimum model. It was previously found that CPM value did not affect the predicting model in the 9 parameters (model A). There was slightly improvement in model (A) during implementing the setting. When the model was tested the result became excellent with error set in the formula.

Figure 9 and 10 represent both the optimum maximum daily temperature (maxdt) and maximum relative humidity (maxrh) respectively for January 1992.

Figure 9. Optimum model performance
for maximum daily temperature

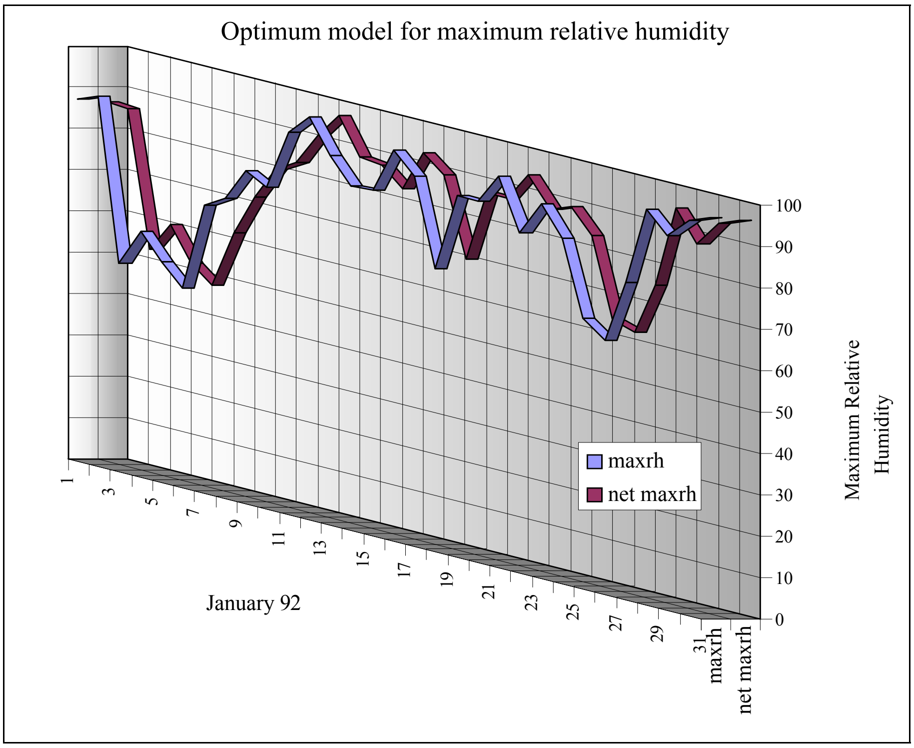
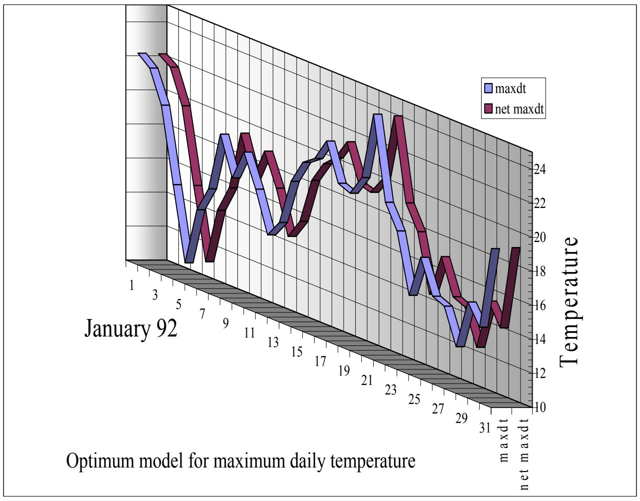


Figure 10. Optimum model performance
for the maximum relative humidity



The results obtained for model (A), was exported to Statistical program (Statistica 1995) to visualize the performance in three dimensional. The desired and the predicted results for both (maxdt and maxrh) are shown in Figures 11 and 12.

FIGURE 11. Surface plot for the performance of model (A).

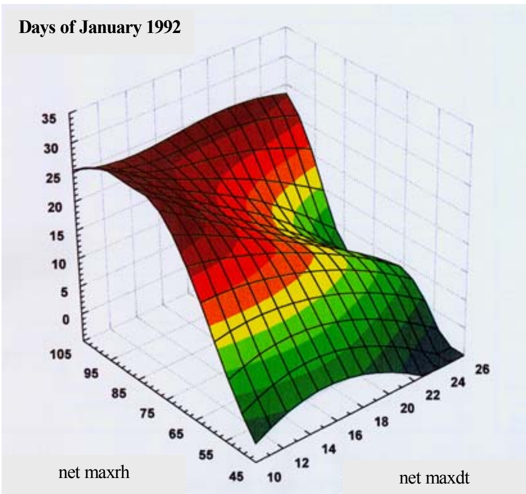
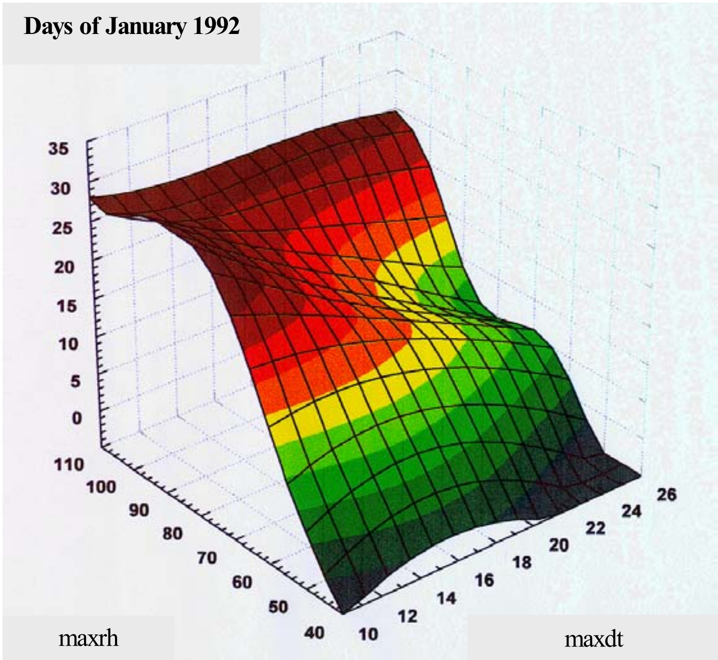


FIGURE 12. Surface plot for the actual data derived from MEPA
for the maximum temperature and maximum relative humidity.



Conclusion and Further Study

SILS provided accurate prediction of the two required values of the weather elements, maximum daily temperature and maximum relative humidity. SILS proved to be a reliable statistical inductive learning network. It was found that optimum model and result could be improved and even the error reduced. Management utilities within the program can manipulate the data to fine-tuning the network performance. The following are some recommended to carries future work:

1. The data to be trained should be a good representation of the problem that need to be solved (i.e. the input variables have relation of output variables.)
2. The data to be trained should be adequate; that the statistical model produced has the ability to construct good automatic function

relationship regarding input-output model.
3. Statistical model improves and works better with large data.
4. Break the problem into different parameters for testing models. Optimum final model should be based on the evaluation of synthesizing specifications (i.e. standard deviation, mean square root, maximum absolute error…. etc).
5. Begin with default synthesize setting as more complex model do not give good result (depending on the data), because sometimes functioning relationship degrading with high complexity of model synthesizing (low value of CPM).
6. Investigate data before modeling using third party graphical applications (ID5, Statistica, Mathcad). It is advised to visualize data in multi-dimensional graphs rather than in tabulated numbers form.
7. Develop the management utilities of the SILS in way that experimenter could export the trained and evaluated data in a third party graphical applications (i.e. Statistica, ID5).
8. Update information to maximum utilization of the predicting model.
9. Further development of study is to develop a model use the extrapolation by predicting temperature and relative humidity for individual days.

Acknowledgments

The author wants to acknowledge the Meteorological of Eastern Province of Saudi Arabia for providing the necessary data to his graduate student Mr. Ali Al-Momen, whom his help is acknowledged. The author also gratefully acknowledges the Fulbright scholarship during his research at the University of Maryland, College Park in the USA during the academic year 1999-2000.

References

1. Abdel-Aal, R. E., Elhadidy, M. A. (1994). A Machine-Learning Approach to Modelling and Forecasting the Minimum Temperature at Dhahran, Saudi Arabia, *Energy*, **19**, (7), 739-749.
2. Barron, A. R., (984). Predicted square error - a criterion for automatic model selection. In S. J. Farlow, (Ed), *Self-organizing methods in modeling: GMDH Type algorithms*, Marcel Dekker, 87-103.
3. IDL™ (Interactive Data Language) (1998). *Using IDL; IDL Version 5.2*, Research Systems Inc.
4. Jadid, M., N., (2000) Displacements analysis by implementing statistical inductive learning and scientific visualization, Accepted for Publication in: The *Journal of the University of Kuwait (Science).*
5. Jadid, M. N. & Fairbairn, D. (1996)**.** Neural network applications in predicting moment-curvature parameters from experimental data. *Engineering Applications of Artificial Intelligence* **9(3)**: 303-319.
6. Jadid, M. N. & Fairbairn, D. (1998)**.** A neural network framework approach to assist experimental investigations. *Emirates Journal for Engineering Research (EJER)* **3(1)**: 13-21.
7. ModelQuest Prospector™ (1997). *User's Manual; ModelQuset Prospector Version 5.0*, AbTech Corporation.
8. Montgomery, G. J., Drake, K. C. (1990). Abductive networks. *Proceedings SPIP* (1294) (Applications of Neural Networks), 56-64.
9. Statistica™ (1995) *User's Manual; Version 5*, StatSoft™ Inc.
10. SubbaNarasimha, P. N., Arinze, B., Anandarajan, M., (2000), "The Predictive Accuracy of Artificial Neural Networks and Multiple Regression in the Case of Skewed Data: Exploration of Some Issues", *Expert Systems Applications*, **19**, pp. 117-1123.

توقع درجة الحرارة والرطوبة النسبية للهواء باستعمال نموذج المحاكاة الإستقرائي الإحصائي

منصور بن ناصر الجديد

قسم علوم وتقنيةا لبناء – كلية العمارة والتخطيط
جامعة الملك فيصل – الدمام
المملكة العربية السعودية

الملخص :

يقدم هذا البحث طريقة لإحتساب توقعات ورؤية وحدة القياسات لدرجة الحرارة اليومية وكذلك درجة الرطوبة النسبية القصوى لشركة كهرباء المنطقة الشرقية في المملكة العربية السعودية عن طريق استعمال نموذج المحاكاة من التعلم الإستقرائي الإحصائي.

تحتوي البيانات على ٣٦ بياناً لحالة الطقس اليومية (وحدة القياس المتري) المتوفرة عن طريق البيانات المدونة من قبل الأرصاد الجوية بالمنطقة الشرقية في المملكة العربية السعودية. وقد تم استعمال برنامج نموذج المحاكاة من التعلم الإستقرائي الإحصائي لإيجاد وتوفير بيانات تلقائية عن طريق التدريب والتقييم حيث ان النتيجة هي الوصول الى علاقة وثيقة بين كل من المدخلات والمخرجات في التوظيف التلقائي مبنياً على النموذج الإحصائي.

لقد ركزت الدراسة على تهيئة نماذج مختلفة كمثل جهاز المضاعف المعقد وعدد الطبقات وحجم الطبقات الأستقرائية التي تعتبر جزءاً من البرنامج. هذا وقد تم تنظيم البيانات لتلائم الاعداد المختلفة لوحدات القياس المترية وعولجت إما عن طريق نماذج التفكيك أو عن طريق نماذج التركيب حتى يتم التوصل الى النموذج الأمثل.

ولقد تم توفير بيانات الطقس اليومية للحد الأقصى لدرجات الحرارة اليومية وكذلك الحد الأقصى للرطوبة النسبية للنموذج التركيبي لاربعة أعوام (١٩٩٠ ، ١٩٩١ ، ١٩٩٣ ، ١٩٩٤) كنوعاً من التدريب.

اما البيانات المطلوبة توقعاتها فهي الحد الأقصى لدرجات الحرارة اليومية والحد الأقصى للرطوبة النسبية من أي يوم من عام ١٩٩٢م. ولقد تم احتساب نسبة الخطىء معتمداً على القيم المتوقعة والقيم الحقيقية.