Journal of Environmental Protection and Ecology 15, No 3, 1101–1109 (2014)

SIMULATING FOR PREDICTING THE HOURLY DEW POINT TEMPERATURE USING ARTIFICIAL NEURAL NETWORKS

C. ILIE[a], M.-L. LUNGU[b]*, L. PANAITESCU[b], M. ILIE[c], D. LUNGU[b], S. NITA[d]

[a]*Faculty of Mechanical, Industrial and Maritime Engineering, Ovidius University of Constanta, 124 Mamaia Blvd., 900 527 Constanta, Romania*
[b]*Faculty of Natural and Agricultural Sciences, Ovidius University of Constanta, 124 Mamaia Blvd., 900 527 Constanta, Romania*
E-mail: *dumilungu@yahoo.com*
[c]*Faculty of Economicl Sciences, Ovidius University of Constanta, Romania*
[d]*USAMV B, Timisoara, Romania*

Abstract. This paper proposes the modelling and simulation of dew point temperature (°C) considering the influence of the following parameters: air temperature, atmospheric pressure, vapour tension and relative air humidity, measured in the following geographic coordinates: Ø 44º31′, λ 29º34′, H 13.5 m, Hb14.0 m, during the year 2011. The importance of this research comes from that a method for predicting the dew point at different times of the day as determined. The dew point is very important in many ways. The dew point determines how you will feel uncomfortable on warm or cold days. It determines whether will it rain or snow. Dew point determines how big to burn hazard vegetation during the dry wind. That is just a few reasons why it is significant to this study. The used method considers the modelling and simulation of the influence that the above-mentioned parameters have over the dew point. Using this method the authors wanted to forecast the value of dew point at a certain time of day, different and after the time of the measured parameters. Those modelling and simulation were performed using a feedforward neural network with the Levenberg-Marquardt training algorithm. Following data analysis, training, testing and validation of the neural network an efficient neural network training was achieved, in conformity with initial established error values and with satisfactory results for validation and testing. Thus, the trained network can predict with a limited error the dew point value at a certain time of the day after the time of measurement parameters that influence the dew point temperature.

Keywords: dew point, artificial neural networks, forecasting, weather.

AIMS AND BACKGROUND

The dew point is the temperature at which the moisture in the air begings to condense into dew or water droplets[1–4]. The accurate estimation of the dew point is very important as it determines whether it will rain or snow as well as how high the danger is for a grass or brush fire during a dry spell[5–8]. It also can be used for

* For correspondence.

determining the amount of the available moisture in the air as well as for estimating the near surface humidity which is crucial from an agricultural viewpoint[9] and tourism activities[10–12].

Meteorological parameters. This studies proposes the modelling and simulation of dew point temperature (°C) (DPT_7 and DPT_13) considering the influence of the following parameters: air temperature (AT_7 and AT_13) – fundamental estate parameter which defines the thermal estate of air, more accurate, the state of thermodynamic equilibrium, atmospheric pressure (AP_7 and AP_13), vapour tension (VT_7 and VT_13) and relative air humidity (RU_7 and RU_13), measured in the following geographic coordinates: Ø 44º31′, λ 29º34′, H 13.5 m, Hb14.0 m, during the year 2011 (every day at 1:10 and 5:00 p.m.).

Artificial Neural Network (*ANN*). The Artificial Neural Networks (ANNs) are used in many fields of research and applications such as: pharmacy, medicine, financial forecast, process and system modelling and simulation, meteorology, speech, handwriting and face recognising, management decision-making, etc.[13,14]

The ANNs are mathematical representations of the biological human brain as proof to use the most powerful processing tool. Hence, the human biological neurons are replaced by artificial ones and the function between them is substituted with mathematical function such as linear, hyperbolic tangent, sigmoid, etc.[15–18]

RESULTS AND DISCUSSION

ANNS TRAINING AND TESTING

For the present research the following software was used: Alyuda NeuroIntelligence and the following machine hardware configuration: Processor – Intel Core2 Duo CPU E8200@ 2.66 GHz, Installed memory (RAM) – 4.00 GB and a 32-bit Operating System.

Step 1 – Initial data analysis. The data consist in values of the five parameters measured daily 2011 in the following geographic coordinates : Ø 44º31′, λ 29º34′, H 13.5 m, Hb14.0 m. The initial data consist in values measured at 1, 7, 13 and 19 h. For the present research were taken into account the values measured at 7 and 13 h.

The following method will aim to predict the value DPT_13 (at 13 h) as influenced by the value of the other four parameters measured at 7 and 13 h and the influence of dew point temperature at 7 h (DPT_7).

Data analysis consists in dividing data into separate columns, defining types of these columns, filling out missing number values, defining the number of categories for categorical columns, etc. The data analysis determined the following results:

- 10 columns and 313 rows were analysed;

• Data partition results: 214 records to training set (68.29%); 49 records to validation set (15.71%) and 49 records to test set (15.71%).

As it can be seen, the database is divided in 3 sets. While the training sets are used only for training, the validation and testing sets are used also for testing.

Step 2 – Data pre-processing. In this step the analysed data are modified in order for an easier use by the ANN in the next step of training. Pre-processing means the modification of the data before it is fed to a neural network.

Pre-processing transforms the data to make them suitable for neural network (for example, scaling and encoding categories into numeric values or binary) and improves the data quality (for example, filtering outliers and approximating missing values), as different software uses different methods[19,20]. The pre-processing used for the present research, considering the type of problem and the previous analysis, but also the researches, the pre-processing type used was One-of N. The One-of-N encoding means that a column with N distinct categories (values) is encoded into a set of N numeric columns, with one column for each category. For example, for the capacity column with values 'Low', 'Medium' and 'High', 'Low' will be represented as {1,0,0}, Medium as {0,1,0}, and High as {0,0,1}. In Table 1 the characteristics of pre-processed data are shown.

Table 1. Results of the pre-processed data pool

Parameter	Scaling range	Min.	Max.	Mean	Standard deviation	Scaling factor
AP_7	[–1..1]	994.2	1032.8	1014.61	7.329456	0.051813
AP_13		993.2	1033	1014.52	7.23138	0.050251
AT_7		–4.4	27.4	12.57	8.210166	0.062893
AT_13		–3.9	29.7	13.57	8.646634	0.059524
VT_7		2.9	28.7	12.79	6.425913	0.077519
VT_13		2.6	30.6	12.81	6.576565	0.071429
RU_7		0	100.0	70.62	24.683013	0.02
RU_13		0	100.0	66.36	24.167374	0.02
DPT_7		–10	23.4	8.77	7.868472	0.05988
DPT_13	[0..1]	–11.4	24.5	8.77	7.932322	0.027855

So the pre-processing process was as following: columns before pre-processing – 10; columns after pre-processing – 10; input columns scaling range – [–1..1]; output column(s) scaling range – [0..1]; numeric columns scaling parameters – AP_7: 0.051813, AP_13: 0.050251, AT_7: 0.062893, AT_13: 0.059524, VT_7: 0.077519, VT_13: 0.071429, RU_7: 0.02, RU_13: 0.02, DPT_7: 0.05988, DPT_13: 0.027855.

Step 3 – Artificial neural network structure. Considering the characteristics of the simulated process, many ANN structure can be determined and compared through the specificity of the process and the data that are belong used for simulations. Thus, the feed-forward artificial neural network is considered the best choice for the present research[21,22].

After building and testing several ANN with feed-forward structures the best ANN network is defined and it has the following structure (Fig. 1): 9 neurons in the input layer, 11 neurons in three hidden layer (first hidden layer – 11 neurons, second hidden layer – 7 neurons and 3 neurons in the third hidden layer) and 1 neuron in output layer.

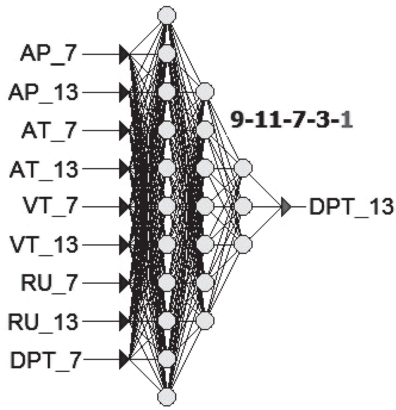

Fig. 1. Simplified graphic representation of feed-forward ANN with structure 9-11-7-3-1 (Source: our own simulation)

This ANN structure is the result of several basic training having the goal to determine the fittest ANN for the database that the researchers used.

Even that it has been stated that in function approximation problems, the hidden layer activation function should be sigmoid-based, and the output layer activation function should be linear-based (the simplest activation function), the researchers choose for the hidden layers the hyperbolic tangent as the activation function (ratio between the hyperbolic sine and the cosine functions (or expanded, as the ratio of the half-difference and half-sum of two exponential functions in the points z and $-z$). This choice is considered because the researchers found that the hyperbolic tangent activation function offers practical advantages of faster convergence during training and that should give a more close value of the training error. The output layer activation function is a logistic function (sigmoid curve)[20]. The researchers consider that the logistic function minimised the effects of over-fitting that happened with the previous trainings when is used the linear function.

Step 4 – Training. Being an essential step in the use of ANN, the training must use certain training algorithms which essentially modify the structural elements of ANN (weights) through several iterations. Those modifications establish the future ANN accuracy. For the selected ANN, the most common training algorithm is Levenberg-Marquardt.

The Levenberg-Marquardt algorithm is an advanced non-linear optimisation algorithm. It is the fastest algorithm available for multi-layer perceptrons (a node or an unit in the interconnected artificial network). However, it has the following restrictions: it can only be used on networks with a single output unit, it can only be used with small networks (a few hundred weights) because its memory requirements are proportional to the square of the number of weights in the network, it is only defined for the sum squared error function and therefore it is only appropriate for regression problems[19].

The stop-training conditions were a maximum absolute training error value of 0.05 tracked only on training set and the local minima avoidance for the Levenberg-Marquardt algorithm was used. The results of training are number of iterations: 297 (time passed: 00:01:40 min). The stop-training reason is: no error improvement. After all iterations were done the absolute training error is 0.019401. The results of training details are presented in Table 2.

Table 2. Training details

Re-train	Iters.	Train error	Validation error	Test error	AIC	Correlation	*R*-squared	Stop reason
1	297	0.052677	0.12265	0.253639	–1334.244	0.999972	0.99994	no error improvement

In Table 2:

– AIC is Akaike information criterion used to compare different networks with different weights (hidden units). With AIC used as fitness criteria during architecture search, simple models are preferred to complex networks if the increased cost of the additional weights (hidden units) in the complex networks do not decrease the network error;

– Iters are the iterations;

– *R*-squared is the statistical ratio that compares model forecasting accuracy with accuracy of the simplest model that just use mean of all target values as the forecast for all records. The closer this ratio to 1 the better the model is. Small positive values near zero indicate poor model. Negative values indicate models that are worse than the simple mean-based model. Do not confuse *R*-squared with *r*-squared that is only a squared correlation[19].

The initial results of the training are shown in Fig. 2 by Dataset errors and in Fig. 3 as the importance considered by the ANN training.

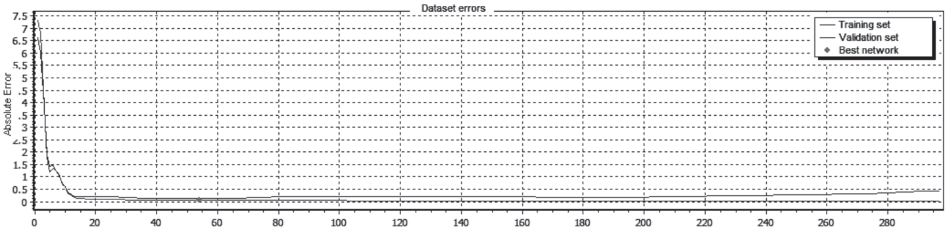

Fig. 2. Dataset errors due ANN training (Source: our own simulation)

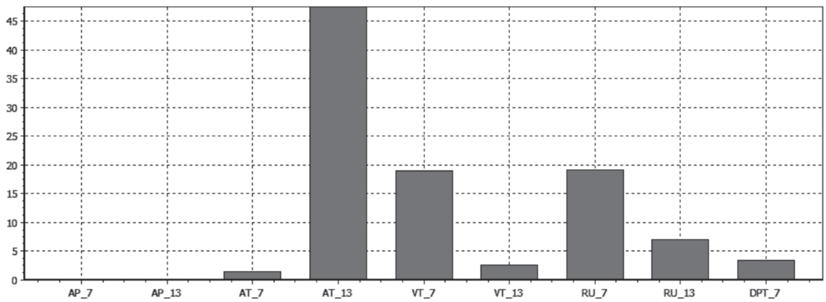

Fig. 3. Input importance over the output (DPT_13) (Source: our own simulation)

Input importance chart shows the most important columns that had the biggest influence on the network. An input column importance is calculated as degradation in network performance after the input was removed and was not used by the network[19]. The input importance over the network training has the following values: AP_7: 5.515543 %, AP_13: 8.596883 %, AT_7: 5.159145 %, AT_13: 3.89001 %, VT_7: 8.254052 %, VT_13: 7.987059 %, RU_7: 30.314 %, RU_13: 30.283306 %, DPT_7: 0 %.

Step 5 – Validation and testing. Validation and testing is a process of estimating quality of the trained neural network. During this process a part of the data that is not used during training is presented to the trained network case by case. Then, the forecasting error is measured in each case and it is used as the estimation of network quality. The automatic testing evaluation is shown in Table 3. Also, differences between the (real) target values and the (simulated) output values of DPT_13 are presented in Fig. 4.

Table 3. Testing summary

Test evaluation	Target	Output	Absolute error (AE)	Absolute relative error (ARE)
Mean	8.771154	8.758222	0.095228	3.36E+12
Std. dev.	7.932322	7.962622	0.339472	1.18E+13
Min.	–11.4	–11.31575	0.000697	0.000117
Max.	24.5	24.419744	5.040020	1.98E+14

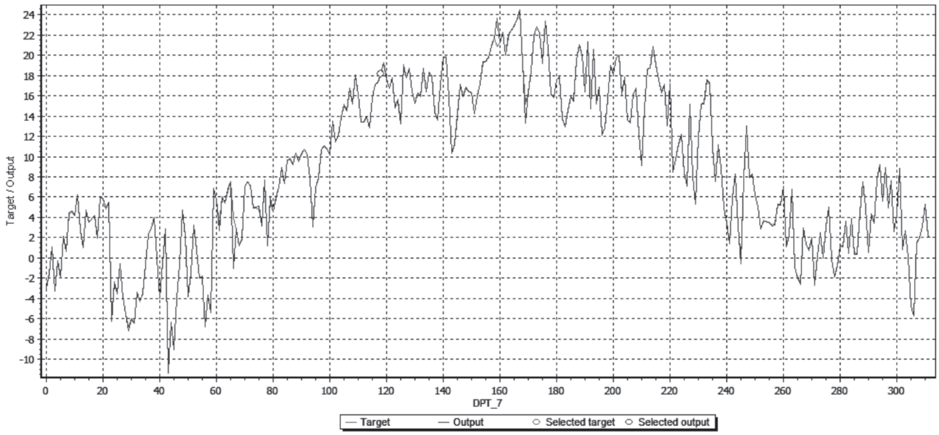

Fig. 4. Actual versus output values comparison (Source: our own simulation)

In Fig. 4 is shown the results of training validation and testing. The data compared are the target (real values) and the output (ANN simulated values). As can be easily seen the error (difference between the real and simulated data) are very small and that means that the initial training is a success and the developed ANN can be used in the next phases of the research.

Free testing. For a better evaluation of the training process another free testing was considered. This test was characterised by the use of new input data that was never fed to the ANN. The importance of this test comes from the wish of the researchers to further more testing the ANN after the automatic testing and validation. This is because a free evaluation of the training should be made. For the free test a number of 21 new input data was fed to the trained ANN. The results of the free test are presented in Fig. 5.

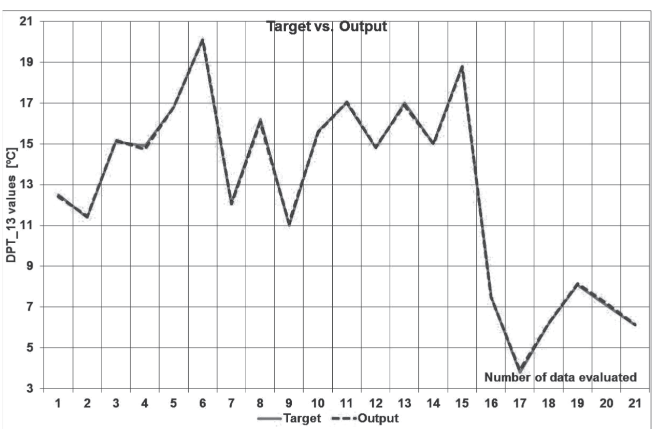

Fig. 5. Free test results: difference between real values for actual and output data

As it can be seen, the results of free testing are almost identical. The differences between the real and simulated data have values from –0.13 and 0.16 as absolute differences and between 2.89 and 1.07 % in relative difference. So the absolute errors are smaller than 0.16 and relative errors are smaller than 2.89%.

CONCLUSIONS

The research delivered the following results:

– Building, training and testing an artificial neural network that can predict the dew point temperature considering several climatic parameters;

– Establishing;

– Considering the validating and testing results, the ANN training is considered a success and all the initial research conditions were achieved. Thus, the ANN can be implemented for the prediction of PDT_13.

The researchers consider that the first phase of the research was accomplished and the next phases, as following, can be initiated:

Establishing the sustainability for future use of ANN for the forecast of the dew point temperature considering more parameters and different timelines.

REFERENCES

1. J. KEYANTASH, J. A. DRACUP: The quantification of Drought: An Evaluation of Drought Indices. Bull Am Meteorol Soc, **83**, 1167 (2002).
2. A. C. MARICA, A. BUSUIOC: The Potential of Climate Change on the Main Components of Water Balance Relating to Maize Crop. Rom J Meteorol, **6** (1–2), 50 (2004).
3. P. SCHWARTZ, D. RANDALL: An Abrupt Climate Change Scenario and Its Implications for United States National Security., www.environmentaldefense. org/documents /3566_AbruptClimateChange.pdf. , 22 p., 2003.
4. T. ZAHARIA, D. MICU, V. TEODOROVA, D. van ELBURG, V. NITA, V. MAXIMOV, M. GOLUMBEANU: The Coherence of the Romanian Marine Protected Areas Network. J Environ Prot Ecol, **11** (1), 199 (2010).
5. E. de MARTONNE: Une nouvelle fonction climatologique: L`indice d`aridité. La Meteorologie, 449 (1926).
6. Ó. A. DO: Regional Drought Analysis and Mitigation Using the SPI. CD-Rom. In: ICID 21st European Regional Conference, Topic 4. 15–19 May, Frankfurt (Oder) and Slubice – Germany and Poland, 2005.
7. K. NIKOLAU, N. PAPADAKIS, S. BASBAS, P. ERMIDIS, P. SPYROGLOU, P. DIMOU: Spatial and Temporal Variation of Air Quality in Greek Cities. J Environ Prot Ecol, **9** (3), 494 (2008).
8. M. LUNGU, Cr. PALTINEANU, I. F. MIHAILESCU: Climatic and Hydrological Risks. University Publ. House, Bucharest, 2008. 193 p.
9. N. CERNESCU: Classification of Soils with Excess Moisture. Research of Soil. Academy Press R. P. R., Bucharest, 1961, 223–250.
10. A. C. MUNTEAN, F. STREMTAN: Tourists Satisfaction with the Tourist Offer of Alba Iulia Historical Fortress. J Environ Prot Ecol, **13** (4), 2445 (2012).

11. D. BEZLOVA, M. DONCHEVA-BONEVA: Protected Areas at the Black Sea Coast as Natural Resources for Development of Ecotourism. J Environ Prot Ecol, **12** (3), 1179 (2011).
12. M. GOLUMBEANU, F. K. VOSNIAKOS, S. AVDIMIOTIS, P. MENTZELOU, C. CONSTANTIN, M. TEOHAREVA: Upgrade Black Sea Scene. Developing a Research Infrastructure for the Black Sea. J Environ Prot Ecol, **11** (3), 1150 (2010).
13. AJITH ABRAHAM, BERNARD de BAETS et al.: Applied Soft Computing Techniques: The Challenge of Complexity. Springer-Verlag, Berlin Heidelberg, 2006.
14. Z. WASZCZYSZYN: Fundamentals of Artificial Neuronal Networks. Institute of Computer Methods in Civil Engineering, 2000.
15. AKIRA HIROSE: Complex-valued Neural Networks: Theories and Applications. World Science Publishing Co Pte. Ltd, Singapore, 2003.
16. A. SHUKLA, R. TIWARI, R. KALA: Towards Hybrid and Adaptive Computing: A Perspective Studies in Computational Intelligence. Springer-Verlag, Berlin Heidelberg, 2010, p. 36.
17. DE XIAOHUI LIU, PAUL R., COHEN PAUL: Advances in Intelligent Data Analysis, Reasoning about Data. In: Proc. of the Second International Symposium, IDA-97, London, UK, August, 1997, p. 58.
18. W. KEITH, A. HIPEL, I. McLEOD: Time Series Modelling of Water Resources and Environmental Systems. Elsevier Science B.V., No 45, 1994, p. 210.
19. Alyuda Neurointelligence – http://www.Alyuda.com/Neural-Networks-Software.htm
20. D. I. NASTAC: Artificial Neural Networks. Advanced Data Processing. Printech Pub., Bucharest, 2000.
21. S. E. FAHLMAN: Faster-learning Variations on Back-propagation: An Empirical Study. Proceedings of 1988 Connectionist Models Summer School, 1998.
22. C. ILIE, M. ILIE, A.-M. TOPALU, L. MELNIC: The Simulation of US Consumer Credit Fluctuation Using Artificial Neural Networks. Communications of the IBIMA (CIBIMA), IBIMA publishing, 2012.

Received 14 March 2014
Revised 11 June 2014