Short-Term Localized External Temperature Prediction for an Intelligent Greenhouse

Péter Eredics
Budapest University of Technology and Economics
2. Magyar tudósok krt.
Budapest, 1117 Hungary

***Abstract* – Climate control for intelligent greenhouses is currently an active field of research. Model based intelligent greenhouse control systems seem to increase the control performance over traditional solutions. This paper focuses on one aspect of the modeling necessary for the intelligent control systems, namely predicting the external temperature, which is closely related to the actual weather conditions. The paper demonstrates the limited performance of uninformed, simple methods for temperature forecasts, and introduces more accurate solutions using information from the problem domain.**

I. Introduction

Weather forecast is one of the hardest prediction problems in applied sciences. Its quality largely depends on the time horizon, geographical location, regional particulars, required accuracy, and adverse effects of missing the forecast, among others. Generally the shorter the prediction span, the more dependent is the forecast on local phenomena and features of the environment.

The topic of our research is an extremely short term air temperature forecast for greenhouses, as the basis for the intelligent greenhouse climate control. Contrary to the usual weekly or daily forecast for larger geographical regions, an hourly air temperature forecast for a small greenhouse, even if it follows long term trends, can involve fluctuations and effects ruled by the local natural and artificial environment features. It is also a reason why the forecasts based on a detailed thermal model is generally infeasible. The paper proposes an easy pragmatic solution to this seemingly impossible problem.

II. Control Problem

Greenhouses have transparent walls and roof and are widely used for vegetable production and growing flowers. Solar radiation is essential for photosynthesis of the plants, and also to keep the inner temperature within an acceptable range. In the cold season a heating system may also be necessary. Contrary in hot weather other actuators, like roof vents, shading systems, exhaust fans or evaporative cooling may be used to avoid overheating.

The primary goal of a greenhouse control system is to create suitable climate for the plants by controlling the actuators of the house cost-effectively.

Traditional greenhouse control

Almost every greenhouse today has a control system using some kind of environment control computer. Yet despite the available computing power control methods did not change much along with the technology. Even today almost all greenhouse control systems work with independent, set-point based PID controllers [1], developed mainly for microcontroller based control systems. These solutions suffer from the missing synchronization of the actuators, and from their dependence on the user to find the appropriate set-points.

Model based, intelligent control

The efficiency of the control can be improved using more advanced models, see e.g. IntelliGrow [2]. There is much of research on complex control models in greenhouse environment, e.g. model predictive control [1]. Such models could provide better control, and they can be used by a forward planning intelligent control system. Such an intelligent control system is currently under development at the Department of Measurement and Information Systems at Budapest University of Technology and Economics. This system makes the necessary control decisions based on the predicted thermal state of the greenhouse for the following 1-4 hours. Using predictions yields better control efficiency because the system can take into account the regular changes of the thermal conditions (caused mainly by the varying amount of solar radiation through the day).

Predicting the thermal state requires an accurate model of the greenhouse and of the external environment. Modeling can be decomposed into separate models of the external environment and the greenhouse, due to their one-way causal relationship: only external environment affects the greenhouse. This paper describes methods modeling the thermal behavior of the external environment.

Modeling alternatives

Theoretically the best, i.e. the most accurate model would be the parametric model based on physical phenomenon and identified from the measurements. It would be however prohibitive in computing time and structural complexity [3]. Costly modeling the external environment could be avoided by using weather forecasts online available on the internet. Several such weather forecasts are available, but none of them can be used for the current purpose, because these forecasts are regional predictions of the long-term weather trends for the next days (not necessarily accurate for the microclimate around the greenhouse and not developed for high prediction accuracy for the first 1-4 hours). Therefore the intelligent system must create its own thermal predictions for its actual surroundings.

Modeling the external temperature could be achieved by using the black-box modeling, e.g. neural networks. Such a solution makes it possible to create the model on site, and to update it during the operation of the control system. The main drawback of black-box modeling is long and resource (measurement) intensive learning time, while the system is not yet ready to create forecasts. This disadvantage can be avoided by using white-box modeling.

In a white-box modeling scenario the model is created in advance of the application using recorded thermal data, and only prediction is done during the operation of the control system. The usability of such a model highly depends on the similarity of data used for creating the model, and the actual data available for creating forecasts. This similarity can be enforced by using data from geographically close locations.

III. Experimental Data and Data Comparison

The experimental greenhouse is situated in Western Hungary. Two private, professional weather monitoring stations are located close to the greenhouse, recording external temperature data in every hour. The stations served the five temperature time series for the experiments presented in this paper (see Table I).

TABLE I
TIME SERIES

Series ID	Start date	Number of records	Length (days)	Location	Notation in figures
T1	2007.04.05	1687	70	Celldömölk (35 km)	———
T2	2007.08.19	1571	65		— —
T3	2007.07.04	1037	43		- - - - - - -
T4	2007.07.02	1085	45	Bozsok (10 km)	—×—
T5	2007.08.18	1591	66		—o—

The time series cover the most important part of the year with respect to the production in the current greenhouse: namely spring and summer.

The prediction techniques detailed in this paper are evaluated using the average absolute error, as in

$$\overline{E} = \frac{1}{P}\sum_{k=1}^{P}\left|T(k) - T_{\Pr ed}(k).\right| \qquad (1)$$

IV. Simple Methods

This section presents some simple solutions for the temperature prediction problem. These solutions work without knowledge about the expected behavior of the modeled external temperature. Therefore on the one hand their accuracy is far from optimal, on the other hand the implementation is simple, and resource requirements are low.

Sample-and-hold prediction

For the short term weather prediction the most widely used method is sample-and-hold approximation [4], shown in

$$T_{\Pr ed}(k+\Delta) = T(k). \qquad (2)$$

Fig. 1 shows the average absolute error of the prediction on various time series for the 1-4 hours prediction time.

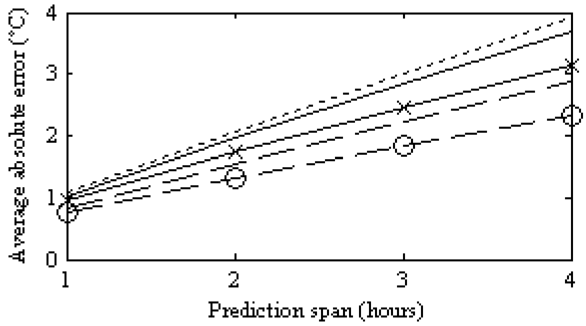

Fig. 1. Average absolute error of the sample-and-hold prediction

In practical applications until recently this accuracy was acceptable, but for the current purpose – to serve as an input for greenhouse modeling and for forward planning – the aggregation of errors calls for better solution.

Linear prediction

The short term accuracy of the prediction can be improved by a linear approximation of the data [5], as in

$$T(k+\Delta) = T(k) + \mu * \Delta * (T(k) - T(k-1)). \qquad (3)$$

The weighting parameter μ limits the influence of the second term of the equation. The optimal value of this parameter was calculated based on the given time series for all four different prediction lengths as, 0.70; 0.63; 0.52 and 0.45 for the 1-4 hours predictions. Fig. 2 shows the average errors of this method using optimal weight settings.

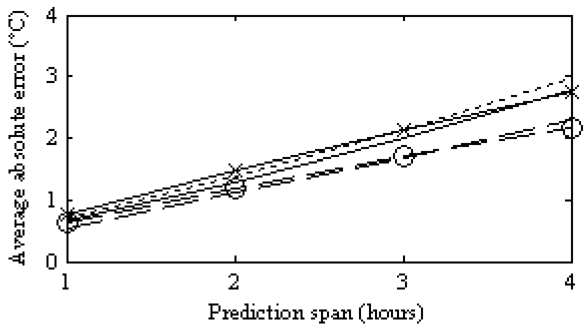

Fig. 2. Average absolute error of the linear prediction

The average error of the linear prediction for one hour span is approximately 0.8 degrees. This error is close to the sensitivity of the sensors used in the control system; therefore further improvements to one hour prediction are not really necessary.

The accuracy of the 4 hours prediction is somewhat below 3 degrees, which is an improvement compared to the sample-and-hold prediction, but not yet accurate enough.

Higher order prediction

To improve the accuracy of the prediction higher order approximations can be used [6]. Equation (4) shows the formula of the second order prediction:

$$T(k+\Delta) = T(k) + \alpha * \Delta * T'(k) + \beta * \Delta^2 * T''(k) \qquad (4)$$

where:

$$T'(k) = T(k) - T(k-1)$$
$$T''(k) = T'(k) - T'(k-1) \qquad (5)$$

Tuning weighting parameters α and β for the 4 hours prediction ended in an unexpected result: the optimal value of alpha was approximately 0.45, while the optimal beta value was 0. This indicates that last term of (4) does not improve the quality of the prediction. The explanation of this effect can be found in the periodicity of the data shown in Fig. 3.

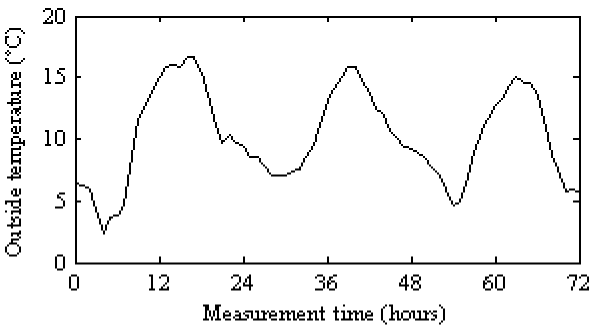

Fig. 3. Periodicity in the temperature data

The calculation of the last term of (4) requires looking back in time for 3 hours. With the prediction length of 4 hours and the sampling frequency of 1 hour the application of the second derivate is based on the assumption that the calculated value is relevant in a 7 hours long time window. The figure above shows that there is an infection point in approximately every 8 hours, therefore the former assumption is almost never true. It means that raising the precision of the 4 hours prediction can not be achieved by using higher order approximations.

V. Informed Methods

Simple methods detailed in the previous section do not use information about the problem domain. Therefore their precision was limited in the longer prediction task. Using information about the phenomena in the background of the measured data can make the prediction better.

Average changes method

The periodicity of the data in Fig. 3 can be easily identified. The change of the external temperature in a given time of the day is very similar to the change on the previous day at the same time. This similarity can be used to create a prediction based on the average change from hour to hour, shown in Fig. 4 for the whole time series.

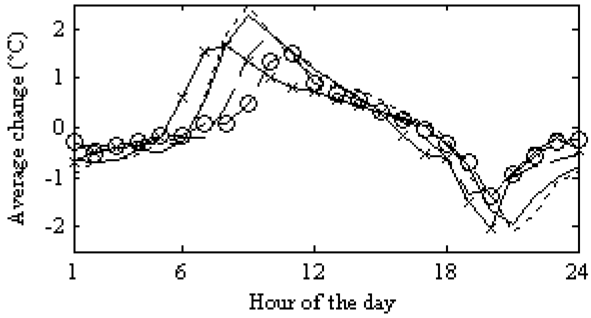

Fig. 4. Average changes of temperature

In practical applications the graphs above are not available for the whole time series. The prediction can only use data from the previous days, shown in (6) below, where P is the number of previous days taken into account:

$$T(k+\Delta) = T(k) + \frac{1}{P}\sum_{i=1}^{P}\big(T(k - i*24) - T(k - i*24 + \Delta)\big) \quad (6)$$

The number of averaged days significant for the prediction can be identified by testing the method with several potential parameters. Fig. 5 shows average absolute error of the method on the T1, T2 and T5 series as a function of the averaging length.

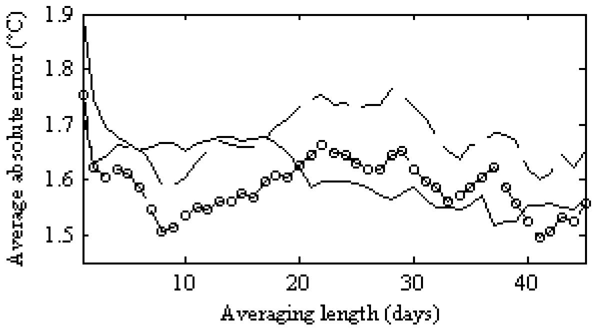

Fig. 5. Average absolute error as a function of averaging length

The T2 and T5 series (both recorded in late summer) have minimum points at 7 days and over 40 days. The explanation of the result is probably the fact, that a 7 days interval describes well the current weather situation, while a 40 days interval extracts the characteristics of the current season. The error using the T1 series (recorded in the spring) drops rapidly in the averaging period of the first 5 days, and remains constant until 20 days. The former explanation can also be used here: 5 days are enough to extract the current characteristics of the weather. Using a longer averaging period makes the approximation smoother, producing slightly better results. Fig. 5 shows average absolute errors in the range of 1.5-1.7 degrees, which is better than the results of all simple methods introduced in the previous sections. Furthermore, this method can be used for longer predictions, with a similar, acceptable average error. Fig. 6 shows the average absolute error as a function of prediction length up to 24 hours.

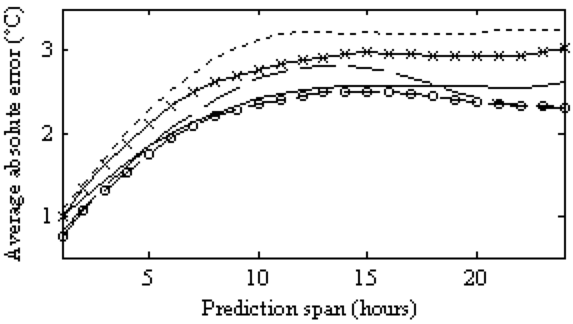

Fig. 6. Average absolute error for longer predictions

For a comparison, the sample-and-hold and the linear approximation both crossed over the absolute error of 3 degrees for the 4 hours length prediction, while the method explained here has this approximate precision for any length less than 24 hours.

Several attempts have been made to further improve the efficiency of this method. Such an attempt was to model the difference in the dynamics of the temperature change using the quotient of the first derivates of the current and the averaged series, as in:

$$T(k+\Delta) = T(k) + \gamma * \frac{1}{P}\sum_{i=1}^{P}\left(T(k-i*24) - T(k-i*24+\Delta)\right). \quad (7)$$

where:

$$\gamma = \begin{cases} \gamma_{\min} & if \;\; \frac{T'(k)}{\overline{T}'(k)} \le \gamma_{\min} \\ \gamma_{\max} & if \;\; \frac{T'(k)}{\overline{T}'(k)} \ge \gamma_{\min} \\ \frac{T'(k)}{\overline{T}'(k)} & otherwise \end{cases} \quad (8)$$

In (7) and (8) P is the number of days taken into account from the past, γ is a weighting parameter calculated as a function of the first derivate of the actual data before the point of prediction and the first derivate of averaged data of the last P days (as defined in (5)). Parameters γ_{min} and γ_{max} are constraints for the value of γ. This modification resulted in 3-7% improved predictions after identification of the γ_{min} and γ_{max} parameters for all series. Unfortunately the parameters had different values for all the time series, and the benefit of the modification was also small, making it practically useless to implement. After some other modification attempts it turned out that a new method had to be found to further improve the prediction precision.

Precedent based prediction

The previously introduced method used the background knowledge, that changes of the temperature are similar on different days at the same time. This approach can be generalized by omitting the time constraints: the precedent based prediction method searches for similar occurrence of the current thermal situation in the past, and uses their subsequent measurements as candidates for prediction. The method is based on the fact that the weather situations are not unique, and similar situations occur usually quite close in time to each other.

The algorithm has several tuning parameters. The span of days (α) has to be specified in which the similar time series parts are searched: there is no point in searching for similar weather conditions in data recorded 6 month earlier. Although the time of the day is not constrained to the hour where the prediction is made, some limitations (β) seem to be useful for the interval of hours searched for similarities: again, there is no reason to search for similar conditions 12 hours earlier.

The current weather conditions are defined by the last few measurements before the prediction. The number of measurements (γ) taken into account for finding similarities is an important parameter. The definition of similarity (err(x_1,x_2)) between the current section of the time series and a candidate time series (the precedent) has to be defined. When calculating the similarities of two series segments to be similar in absolute value is perhaps too strong requirement: a set amount of offset difference should be tolerated, but the maximum value of such offset difference must be limited (δ). The last aspect of similarity to be taken into account is the significance of the location in time of the difference between current time series segment and the candidate. Differences closer in time to the present moment, are more notable for the similarity, and should be punished by a higher weighting parameter (ε^i).

The algorithm calculates the amount of similarity of the possible candidates (all γ-long time series segments within the former α days within the number of β hours, with less offset difference than δ, identified by their start index *l*) by calculating the error value defined as:

$$E(k,l) = \sum_{i=0}^{\gamma-1} err\left(T(k-i), T(l-i)\right) * \varepsilon^i. \quad (9)$$

In (9) the index *k* represents the present, where the prediction has to be made. After evaluating *E(k,l)* for all possible candidates it seem to be straightforward to choose the lowest *E(k,l)* value, and use *T(l+Δ)* as the prediction. This solution would be acceptable only in case of a deterministic modeling problem. In the current situation, where due to missing information it is better to consider the phenomena as random, better results can be achieved by using averaging over the predictions of the best candidates. The number of best candidates used for synthesizing the resulting prediction is the last parameter of the method.

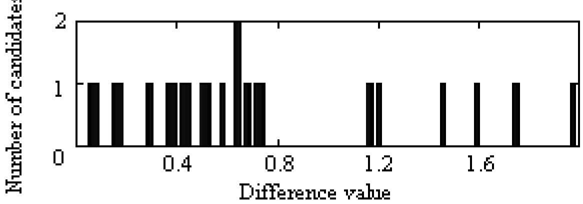

Fig. 7. Number of candidates as a function of *E(k,l)*

In an optimal situation the good and bad candidates should be visually separated, as show in a particular situation in Fig 7. In the figure above the group of candidates with lover *E(k,l)* value can be easily separated from other candidates. Unfortunately using only these candidates for the prediction yields worse results. In addition in most cases the candidates can not be separated easily. In such a situation number of candidates should be fixed by a last parameter (η) of the algorithm. The final prediction of the model is calculated by

$$T(k+\Delta) = \frac{1}{\eta}\sum_{i=1}^{\eta} T(C(i)+\Delta). \tag{10}$$

Where *C(i)* is the start time of candidate *i*. The parameters of the algorithm have been optimized by using all the time series. Table II shows the results.

TABLE II
OPTIMAL PARAMTERS OF THE PRECEDENCE BASED METHOD

Number of days	α	20
Number of hours	β	1
Time series length	γ	3
Error function	$err(x_1,x_2)$	$(x_1-x_2)^2$
Maximum offset difference	δ	7
Difference weight over time	ε	0.8
Number of best candidates used	η	18

Using the best parameter set, the algorithm computed the prediction shown in Fig. 8. The accuracy of the prediction for the time series was between 1.5-1.7 degrees.

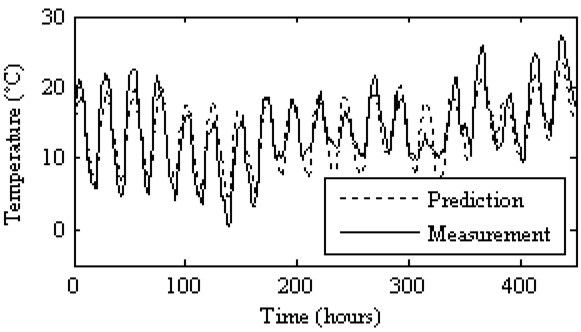

Fig. 8. Results of the precedence based prediction

The total accuracy of the method was even better by 0.2 degrees, when single *Δ* step prediction was decomposed into *Δ* one-step predictions. Unfortunately further attempts to lower the error by using the former prediction error values; weighting the candidates with their similarity value; or by estimating the prediction accuracy based on the variance of the candidates failed.

The precedence based prediction method has the final accuracy of 1.3-1.5 degrees for the 4 hours prediction.

VI. CONCLUSION

In the paper we show that simple methods were able to predict the external temperature with an acceptable accuracy of 0.8 degree for the oncoming 1-2 hours without incorporating any knowledge about the problem domain. For 3-4 hours long predictions the simple methods produced average absolute error of more than 2 degrees, unsuitable for the intelligent control purpose, and calling for more informed solutions.

Using the average change of the temperature from an hour to an hour could be used to provide 1.5-1.7 degree average absolute error for the 4 hours long prediction. Even lower, 1.3-1.5 degree average absolute error is realized using the precedence based prediction method. In this case the basic assumption is the high probability of similar weather conditions close in time to each other. Finding similar conditions in the past, and using their subsequent segments as candidates for the prediction yielded the lowest absolute average error among all examined methods.

We have shown thus that simple temperature prediction methods mining in the past weather data records (geographically even loosely related to the investigated localization) produce predictions of accuracy satisfactory for the further development of intelligent control solutions. The gain is clearly that we can dispense with costly (up to be infeasible, or impossible) parametric, or black-box solutions. As an intelligent greenhouse collects its own climate data, with time weather records from weather station localized strictly by the greenhouse can be mined acc. to the algorithm, increasing possibly further the prediction accuracy.

REFERENCES

[1] M.Y. El Ghoumari, H.-J. Tantau, J. Serrano. Non-linear constrained MPC: Real-time implementation of greenhouse air temperature control. Computers and Electronics in Agriculture, 49: 345-356, 2005.
[2] J.M. Aaslyng, J.B. Lund, N. Ehler, E. Rosenqvist. IntelliGrow: a greenhouse component-based climate control system. Environment Modelling & Software, 18: 657-666, 2003.
[3] T. Kolarik, G. Rudorfer. Time series forecasting using neural networks. ACM Press, Proceedings of the International Conference on APL, 86-94, 1994
[4] John Rowland Higgins. Sampling theory in Fourier and signal analysis: foundations. Oxford University Press, 1996.
[5] A. D .Papalexopoulos, T. C. Hesterberg. A regression-based approach to short-term system load forecasting. IEEE Transactions on Power Systems, 4: 1535-1547, 1990
[6] M. C. Hubbarda and W. G. Cobourna. Development of a regression model to forecast ground-level ozone concentration in Louisville, KY. Atmospheric Environment, 14-15: 2637-2647, 1998