

Estimating Missing Weather Data for Agricultural Simulations Using Group Method of Data Handling

M. C. ACOCK

*Remote Sensing and Modeling Laboratory, U.S. Department of Agriculture Agricultural Research Service,
Beltsville, Maryland*

YA. A. PACHEPSKY

*Remote Sensing and Modeling Laboratory, U.S. Department of Agriculture Agricultural Research Service,
Beltsville, Maryland, and Duke University Phytotron, Duke University, Durham, North Carolina*

(Manuscript received 28 April 1999, in final form 12 July 1999)

ABSTRACT

Contacting weather stations via modems to obtain weather data for crop simulations has become a common practice. Users sometimes encounter gaps in these data, and techniques are needed to estimate weather variables for days when the data are absent. The authors hypothesized that such estimations can be made using data from before and after the day with no data. Dependencies of weather variables of a particular day on weather variables from several days before and after could be very complex. To find and to express these dependencies, group method of data handling (GMDH), which is a tool for modeling complex "input-output" relationships by building hierarchical polynomial regression networks, was used. Data on daily solar radiation, maximum and minimum temperatures, and wind runs collected daily in Stoneville, Mississippi, during May–September of 1982–92 were used. Fourteen-hundred sequential 7-day datasets from the database were extracted. For each dataset, the authors assumed that weather variables on the fourth day were unknown and had to be found from the weather variables of days 1, 2, 3, 5, 6, and 7. Seventy-five percent of these data were used to find the hierarchical polynomial regression, and 25% were used to evaluate it. Correlation coefficients between calculated and actual parameters were similar for training and evaluation datasets. Coefficients of determination (R^2) were about 0.88 for minimum temperature, 0.80 for maximum temperature, and 0.80 for wind run. Accuracy of the solar radiation and precipitation estimates was lower, and R^2 was about 0.2–0.3 but improved to 0.5–0.6 for the training dataset and 0.3 for the validation dataset for both variables when an additional indicator variable that shows the presence or absence of rain was included. The next day after the day with missing data gave the most essential information. Increasing the number of missing days resulted in gradual deterioration of the accuracy for all variables but wind run. GMDH can be a useful tool for filling gaps in weather data from weather stations installed in the field.

1. Introduction

Weather data from automated weather stations have become an important component for prediction and decision making in agriculture and forestry. The data collected from such stations are used in predictions of insect and disease damage in crops, orchards, turfgrasses, and forests (Dinelli 1995); in deciding on crop-management actions such as irrigation (Parker and Zilberman 1994); in estimating the probability of occurrence of forest fires (Fujioka 1995); and in many other applications. Both networks of weather stations and in-

dividual stations accessed via modem currently are in use (Snyder et al. 1996).

The reliability of the automated weather stations has been improved greatly (Elwell and Klink 1993). Failures of weather stations still happen (Mott et al. 1994), however, and techniques are necessary to fill gaps in data sequences to use the data as input in agricultural models. Several approaches have been pursued by researchers. Mott et al. (1994) proposed developing regression equations from data obtained from neighboring weather stations. Ashraf et al. (1997) used geostatistics to interrelate data from several neighboring weather stations.

If the density of the weather stations is low, the use of neighboring stations becomes problematic because daily weather patterns often are strongly related to local landscape conditions (Blackie and Simpson 1993). Therefore, using local historical weather records may

Corresponding author address: Ya. A. Pachepsky, Remote Sensing and Modeling Laboratory, Bldg. 007, Rm. 008, BARC-WEST, Beltsville, MD 20705.
E-mail: ypachepsky@asrr.ars.usda.gov

be a viable alternative to using neighboring stations in reconstructing missing data.

The reconstruction of missing weather data is different from weather forecasting because both data collected before the gap and data collected after the gap can be used. Still, the dependencies between the missing and available weather variables are expected to be complex, and advanced data analysis tools are needed to find and to express these dependencies with sufficient accuracy.

Group method of data handling (GMDH) is a term for the group of algorithms developed to simulate observed complex "input-output" dependencies. GMDH provides an automated selection of essential input variables and builds hierarchical polynomial regressions of necessary complexity (Farrow 1984). GMDH algorithms were applied successfully to predict seasonal precipitation (Lebow et al. 1984), monthly mean temperature (Lin et al. 1994; Setoodehnia et al. 1993), and daily minimum and maximum temperature (Abdel-Aal and Elhadidy 1994, 1995). GMDH algorithms are self organizing so that they do not presume any fixed functional dependence between inputs and outputs; this dependence evolves as the modeling with GMDH proceeds.

The objective of this paper was to explore the applicability of GMDH to reconstruct missing daily weather variables when preceding and subsequent daily weather variables were available. Historical weather records for a particular location in Mississippi were used to build and to test GMDH models. The location represents the region for which the so-called GLYCIM crop simulator is used on farms having weather stations (Reddy et al. 1995) at which gaps in data sometimes occur.

2. Materials and methods

a. Design of the study

We estimated key daily weather variables from data collected 3 days before and 3 days after the day in question. Fourteen-hundred sequential groups of 7-day daily weather data were extracted from weather records collected during May–September in Stoneville, Mississippi, for 10 consecutive years. We included data for the May–October period that covered the growing season for soybean and cotton crops. The daily weather variables were solar radiation (Rad), maximum and minimum temperature (Tmax and Tmin), precipitation (Prec), and wind run (WndR).

For each 7-day group, we assumed that weather variables on the 4th day were unknown and had to be found from the weather variables of days 1, 2, 3, 5, 6, and 7, and from the day of the year of the fourth day (IDAY). Below, the day number in the 7-day group is attached to the symbol of the weather variable so that Rad5 denotes the solar radiation for the 5th day of the group and the solar radiation for the day after the missing data day, Tmin2 denotes the minimum temperature for the

2d day of the 7-day group and the minimum temperature two days before the missing data day, and so on.

We also wanted to see how the accuracy of estimates will change when we use different numbers of days after the missing data day. For that purpose, we used the same 7-day groups and repeated estimations given the assumption that the weather variables were unknown on the 5th, 6th, and 7th days. When the weather variables were unknown on the 7th day, the estimates were actually representing a weather forecast based on the weather of the preceding week.

There may be cases when data are missing for more than 1 day. We built estimates of average weather variables for the sequences of 2 and 3 missing days. Data for these estimates were extracted from the same May–September weather records collected in Stoneville, Mississippi. We extracted all sequential 8-day groups and 9-day groups and computed average weather variables for days 4 and 5 in 8-day groups and for days 4, 5, and 6 in 9-day groups. Then we assumed that these averages are unknown and have to be found from the weather variables of days 1, 2, 3, 6, 7 and 8 and days 1, 2, 3, 7, 8 and 9 in 8-day groups and 9-day groups, respectively.

b. GMDH algorithm

The idea of GMDH is to employ estimates of the output variable obtained from simple regression equations that include small subsets of input variables (Farrow 1984). Although the accuracy of such estimates is low, these estimates may be better predictors of the output variable than are some of the input variables. The best of these estimates are included in the set of input variables, and, again, small subsets of variables from this set are used to build new estimates.

The general functioning of the GMDH algorithm can be understood from the following example. Let the original data consist of the column of the observed values of y and of N columns of the observed values of the independent variables x_1, x_2, \dots, x_N . The primeval equations are quadratic polynomials:

$$z = A + Bx_j + Cx_k + Dx_j^2 + Ex_k^2 + Fx_jx_k, \quad (1)$$

Here A, B, C, D, E , and F are parameters, and x_j and x_k are two independent variables taken from the total list of independent variables. Each iteration consists of three steps. Step 1 consists of obtaining estimates of y using primeval equations. All independent variables x_1, x_2, \dots, x_N are taken two at a time to become x_j and x_k in Eq. (1) and regression polynomials [Eq. (1)] are constructed so that values of z best fit the dependent variable y . The total number of these polynomials is $N(N - 1)/2$. The resulting columns of z_m values, $m = 1, 2, \dots, N(N - 1)/2$, contain estimates of y from each polynomial and are interpreted as new improved variables that may have better predictive power than the original $x_1, x_2,$

\dots, x_N . The objective of the next step is to keep only the best of these new variables.

Step 2 consists in screening out the least-effective new variables. There are several selection criteria (Farfrow 1984), all based on mean square absolute or relative error of z_m with respect to measured y , and often including a correction that "punishes" a network for excessive complexity. In some GMDH versions, columns z_m are retained that have the criterion value smaller than a prescribed value. In other versions, a prescribed number of the best z_m are retained.

The list of input variables is modified at the end of this step. In some versions of the method, columns of x_1, x_2, \dots, x_N are merely replaced with the retained columns of z_1, z_2, \dots, z_K , where K is the total number of retained columns. In other versions, the best K retained columns are added to columns x_1, x_2, \dots, x_N to form a new set of input variables. The total number N of input variables changes to reflect the addition of z_m values or the replacement of old columns x_N with z_m .

Step 3 tests whether the set of equations can be improved further. The smallest value of the selection criterion obtained at this iteration is compared with the smallest value obtained for a previous iteration. If an improvement is achieved, steps 1 and 2 are repeated, otherwise the iterations stop, and the network is built.

We have worked with a version of the algorithm coded in the commercial software package ModelQuest (AbTech Corp., 1992–96). In this version, singlets, doublets, and triplets are used as primeval equations.

$$\text{Singlet: } z = \sum_{i<15} a_i x_i.$$

$$\text{Doublet: } z = \sum_{i+j<4} x_i^i x_j^j.$$

$$\text{Triplet: } z = \sum_{i+j+k<4} x_i^i x_j^j x_k^k.$$

One, two, or three independent variables at a time are used to build the primeval equations. A stepwise regression is employed to diminish, if possible, both the number of coefficients and the number of independent variables in the primeval equations. The number of variables retained in the input list is limited to 15. The criterion of Barron (1984) is used to screen out the new variables and to stop the iterations when the complexity is balanced with the accuracy. The whole dataset is divided into training and evaluation datasets in a 3:1 ratio. The training dataset is used to build a hierarchical polynomial network and to estimate its accuracy, whereas the evaluation dataset is used to estimate the reliability of the network. Both the original input variables x_1, x_2, \dots, x_N and the output variable y are normalized to have a mean value of zero and a variance of one, and the normalized variables participate in network building. After the network is built, the results are denormalized to the original values.

An example of the GMDH network estimating daily minimum temperature at day 4 ($T_{\min 4}$) is shown in Fig.

1. The original list of input variables included Rad1, Rad2, Rad3, Rad5, Rad6, Rad7, Tmax1, Tmax2, Tmax3, Tmax5, Tmax6, Tmax7, Tmin1, Tmin2, Tmin3, Tmin5, Tmin6, Tmin7, Prec1, Prec2, Prec3, Prec5, Prec6, Prec7, WndR1, WndR2, WndR3, WndR5, WndR6, WndR7, and IDAY for a total of 36 variables. All of them were normalized. Only three variables, Rad5, Tmin3, and Tmin5, were selected at the first iteration, and a new variable z_1 was retained. At the second iteration, Rad6 and Tmin2 were selected as additional variables that formed a primeval equation with the z_1 variable, and a new variable z_2 was retained. At the third iteration, Tmax3 and WndR5 were selected as additional variables that formed a primeval equation with the z_2 variable, and a new variable z_3 was retained. Last, at the fourth iteration, WndR2 and WndR7 were selected to be coupled with the variable z_3 , and new variable z_4 was retained. After the fourth iteration, the number of coefficients in the model was balanced with the model's accuracy, and the iterations were stopped. The variable z_4 was denormalized to calculate $T_{\min 4}$ in its original range. All primeval equations were cubic polynomials.

Performance of the GMDH method has been compared with "climatology" and persistence estimates. Here climatology is defined to mean that estimates of a weather variable for a particular day of a particular year were obtained as average values for the same day across all years in the database except for the year for which the estimate was sought. The persistence estimates of a weather variable were obtained as values of the same variable measured the day before the day for which the estimate was sought. A statistical comparison of accuracies of the methods was made using F ratios calculated as

$$F_{\text{Clim}} = \frac{s_{\text{Clim}}^2}{s_{\text{GMDH}}^2} \quad \text{and} \quad F_{\text{Pers}} = \frac{s_{\text{Pers}}^2}{s_{\text{GMDH}}^2},$$

where s_{GMDH}^2 , s_{Clim}^2 , and s_{Pers}^2 are variances of errors in estimates obtained with GMDH in validation datasets, climatology, and persistence, respectively. The F ratios were compared with F statistics $F_{N-1,M-1}$, where M is the number of data points in validation datasets, $M = 360$, and N is the total number of data points in the database, $N = 1440$. The F statistics with large numbers of degrees of freedom are not tabulated. We calculated $F_{1439,359}$, as described by Press et al. (1992, p. 222), using the BETAI subroutine published by these authors. Critical values of $F_{1439,359}$ at 0.01 and 0.001 significance levels were 1.33 and 1.15, respectively.

3. Results

The accuracy and reliability of the polynomial networks are shown in Fig. 2 and Table 1. The best results were obtained for minimum temperatures (Fig. 2). About 88% of variation in T_{\min} values could be explained using observed weather variables from 3 days

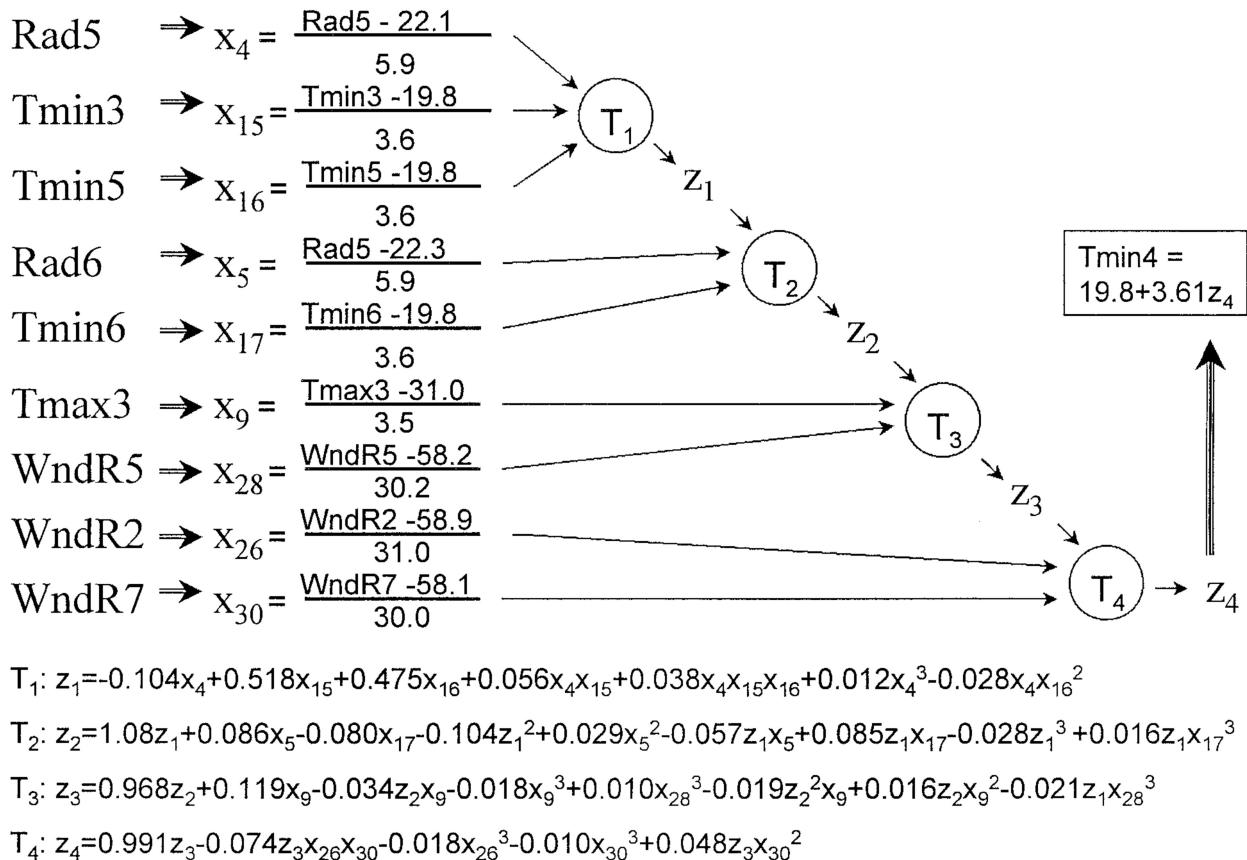


FIG. 1. Structure and content of the GMDH network built to estimate daily minimum temperature at day 4 (Tmax4) from the values of daily solar radiation Rad, daily minimum temperature Tmin, daily maximum temperature Tmax, and daily wind run WndR. See comments in the text.

before and 3 days after the day in question. Root-mean-square error of the estimates was about 1.3°C. The slope of the regressions of estimated minimum temperatures on measured ones was significantly less than 1. Hierarchical regressions built with GMDH tended to overestimate low values and underestimate high values of daily minimum temperature.

Data on daily maximum temperature and daily wind run shown in Fig. 2 demonstrate similar accuracy levels. The percentage of the variation explained by the GMDH hierarchical regression is about 80%. Root-mean-square error of the maximum temperature and wind run estimates is about 1.6°C and 13 km, respectively. High values are underestimated and low values are overestimated.

Estimates of daily radiation and precipitation had relatively low accuracy. The coefficient of determination (R^2) was significant but small, in the range 0.2–0.3. Somewhat better results were obtained when we introduced an indicator variable IP. This variable was set to zero when there was no precipitation on day 4 and was set to one when any precipitation occurred on day 4. With this variable, about 50% of variation in the daily solar radiation and in precipitation could be explained

using weather variables from previous and following days (Fig. 2). The evaluation dataset had values of R^2 near 0.3, significantly less than the value of 0.50 obtained for the training data set. Mean square errors were similar for training and evaluation datasets (Table 1).

Polynomial networks to estimate missing weather variables are shown in Fig. 3. All of the networks have four layers. Table 2 shows how the accuracy of the network progressed as more layers were added. The main improvement in estimation of Tmin4, Tmax4, and WndR4 occurred after the first triplet was found. Therefore, variables included in this triplet were expected to be the most influential predictors. Results from sensitivity analysis (not shown) concurred with this hypothesis. In all cases, the most influential predictors were values of the same weather variables measured the day before and the day after the day of missing datum. The third influential variable was the next-day solar radiation for the minimum temperature, the precipitation on the day before for the maximum temperature, and the next-day maximum temperature for the wind run. The radiation and precipitation estimates gradually improved as new layers were included in the network (Table 2),

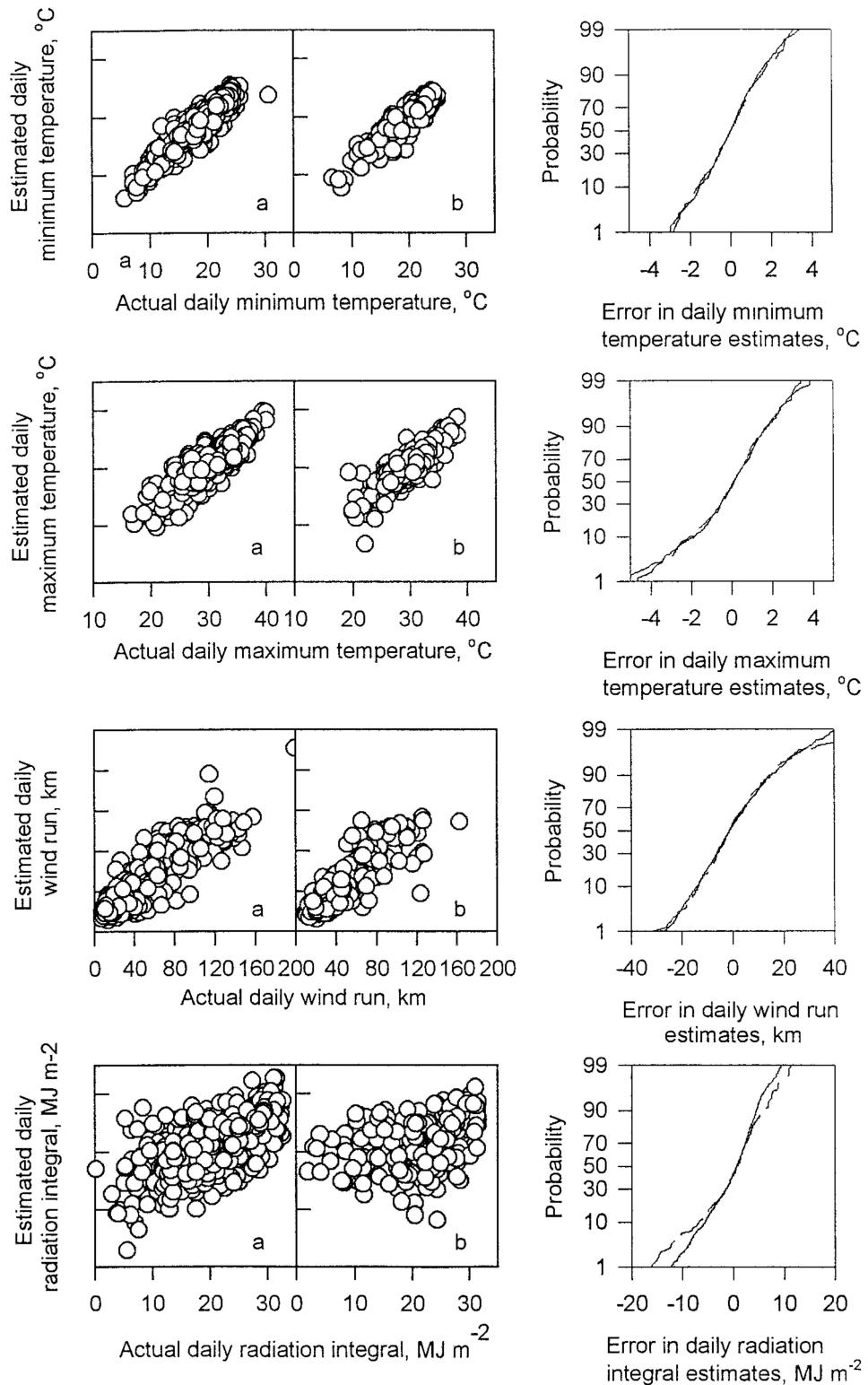


FIG. 2. Plots of 1:1 ratio diagrams and error distributions for GMDH estimates of the daily weather variables: (a) and solid line—training dataset; (b) and broken line—validation dataset.

TABLE 1. Accuracy of the GMDH networks to estimate missing daily weather variables as compared with climatology and persistence.

Variable	Average absolute error		Root-mean-square error				R^2			
	GMDH		GMDH		GMDH		Climatology		Persistence	
	a*	b**	Climatology	Persistence	a	b	a	b	a	b
Min temperature, °C	0.85	0.96	2.38	1.45	1.24	1.25	3.11	2.11	0.88	0.88
Max temperature, °C	1.17	1.18	2.48	1.72	1.60	1.67	3.14	2.44	0.79	0.77
Daily wind run, km	9.3	9.6	28.1	12.5	12.8	13.6	32.6	17.6	0.83	0.79
Daily radiation integral, MJ m ⁻²	3.3	3.7	4.82	4.84	4.4	4.9	6.1	5.9	0.49	0.30
Daily precipitation, mm	2.1	3.2	4.45	4.20	5.2	6.5	9.0	11.2	0.47	0.31

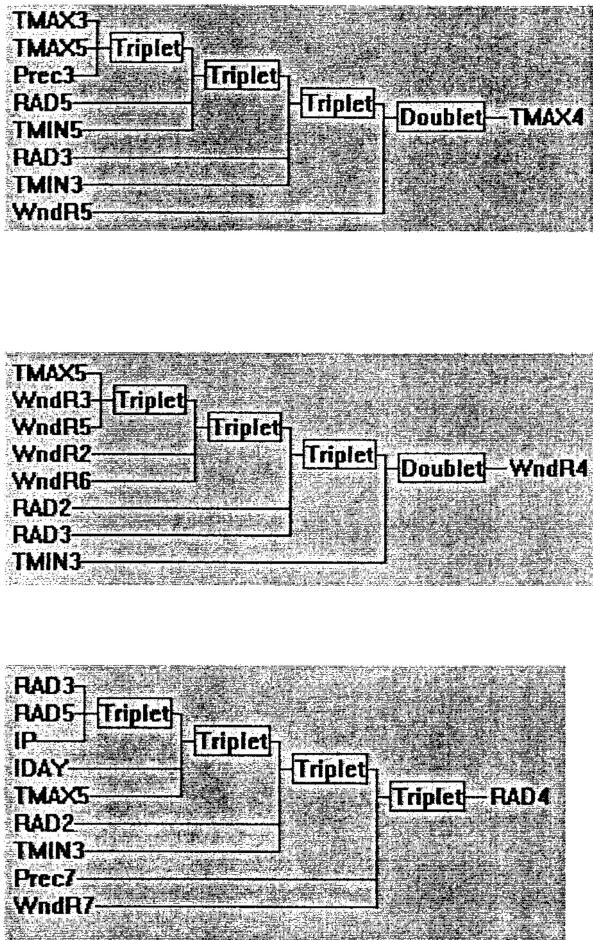
* Training dataset.
** Evaluation dataset.

FIG. 3. Structures of the GMDH networks built to estimate daily maximum temperature, wind run, and radiation integral.

and no single variable except for the indicator variable could be distinguished as the most influential.

Statistics that show the accuracy of the GMDH estimates are compared in Table 1 with the same statistics for climatology and persistence estimates. The climatology estimates are much less accurate than the GMDH estimates in evaluation datasets. For example, root-mean-square errors of maximum temperature, minimum temperature, and wind run estimates are approximately 2 times larger in climatology estimates than in GMDH

TABLE 2. Correlation coefficients between measured and GMDH-estimated weather variables as dependent on the number of layers in GMDH networks.

Variable	1 layer	2 layers	3 layers	4 layers
Min temperature, °C	0.82	0.85	0.87	0.88
Max temperature, °C	0.74	0.76	0.77	0.79
Daily wind run, km	0.80	0.81	0.82	0.83
Daily radiation integral, MJ m ⁻²	0.36	0.41	0.45	0.49
Daily precipitation, mm	0.35	0.44	0.46	0.49

estimates. Differences in accuracy between climatology and GMDH estimates of daily radiation integrals and precipitation also are large. Correlation between measured and estimated-from-climatology weather variables is very weak. The ratios of variances of climatology and GMDH estimation errors exceed the critical value F of 1.33 for $p < 0.001$ (Table 1). Therefore, greater than 0.999 probability exists that the errors in climatology estimates have larger spread than do the errors in GMDH estimates.

Weather variables estimated from persistence are closer to the actual values than those estimated from climatology. All persistence estimates trail the GMDH estimates in accuracy, however. For temperature, radiation integral, and precipitation, the ratios of variances of persistence and GMDH estimation errors exceed the critical value F of 1.15 found for $p < 0.01$ (Table 1). Therefore, greater than 0.99 probability exists that the errors in persistence estimates have a larger spread than do the errors in GMDH estimates.

Table 3 shows changes in the estimation accuracy as related to number of days after the missing-data day. Comparing data in this table with data in Table 1, one can conclude that there is a very small loss in accuracy as the number of days after the missing-day data decreases from 3 to 1. The accuracy of estimates, however, becomes significantly worse when no data after the missing data day are used, that is, when the estimates represent a pure forecast. Using at least 1 day of data after the missing day is essential in filling gaps in weather data.

Results of estimating average weather variables for two and three missing days are shown in Table 4. The accuracy of the temperature, daily solar radiation, and daily precipitation estimates decreases as the length of the gap in data increases. By contrast, the accuracy of the wind-run estimates increases as the number of missing days increases. The deterioration of temperature estimates is not as significant as that for radiation and precipitation.

4. Discussion

The performance of the GMDH model developed to predict a missing daily weather variable depended on the variable. Best results were obtained for maximum and minimum temperatures and for wind run. Therefore, the best use of GMDH network estimates (with the inputs available for this study) can be made in models for which temperatures and wind run are the most important environmental inputs, as in, for example, insect population models. The lack of accuracy in precipitation estimates makes our GMDH networks not suitable to be used to make agricultural management decisions such as irrigation scheduling that are based on the amount of precipitation. This conclusion may be specific for both the region of study and the list of predictors, however. The successful use of self-learning algorithms to predict

TABLE 3. Accuracy of missing daily weather variables estimates as affected by how many days after the day with missing data were used in the estimation.

Daily weather variable	Average absolute error						Root-mean-square error					
	2 days		1 day		0 days		2 days		1 day		0 days	
	a*	b**	a	b	a	b	a	b	a	b	a	b
Min temperature, °C	0.94	0.98	0.95	0.98	1.29	1.36	1.28	1.30	1.27	1.35	1.84	1.88
Max temperature, °C	1.15	1.21	1.17	1.24	1.44	1.46	1.59	1.63	1.59	1.69	1.95	1.96
Wind run, km	9.3	9.7	9.3	9.6	10.6	11.0	12.5	13.9	12.6	12.6	14.3	15.0
Daily solar radiation, MJ m ⁻²	3.16	3.8	3.25	3.97	3.52	3.54	4.26	5.13	4.34	5.28	4.59	4.87
Daily precipitation, mm	2.04	2.90	2.01	2.05	1.89	3.65	5.2	9.46	5.2	6.87	4.92	11.2

* Training dataset.

** Evaluation dataset.

TABLE 4. Effect of the number of missing days on the accuracy of estimating average weather variables for these days.

Variable	Average absolute error						R^2	
	2 days		3 days		Root-mean-square error			
	a	b	a	b	a	b		
Min temperature, °C	1.02	1.08	1.13	1.14	1.39	1.50	1.65	
Max temperature, °C	1.14	1.25	1.21	1.33	1.50	1.64	1.69	
Wind run, km	8.8	9.0	8.55	7.9	11.8	12.0	11.3	
Daily solar radiation, MJ m ⁻²	2.80	2.95	2.75	3.54	3.80	2.98	3.60	
Daily precipitation, mm	2.00	2.66	2.18	2.65	4.15	5.90	3.84	

* Training dataset.
** Evaluation dataset.

precipitation has been reported for conditions in Germany (Dumais and Young 1995).

GMDH models tended to overestimate observations with low values and to underestimate observations with high values. This result may be related to the fact that low-order polynomials are used as regression functions in elements of the networks. These functions are not suited to express rapid changes in output variables.

We compared our results with results of Abdel-Aal and Elhadidy (1995), who used GMDH for weather forecasts in a location in Saudi Arabia. Their mean absolute errors of minimum and maximum temperature were 0.79° and 0.97°C, respectively. These values were lower compared with our values of 0.96° and 1.18°C. The reason for this difference could be that Abdel-Aal and Elhadidy (1995) used 18 weather variables in their predictive models, whereas we worked with five. Another reason may be that the weather patterns are more stable in the Dhahran location studied by Abdel-Aal and Elhadidy than in the region around Stoneville, Mississippi. The accuracy of the minimum temperature estimates was higher than that of the maximum temperature estimates in both our and the Abdel-Aal and Elhadidy studies.

The day of year was included in the GMDH network to improve the daily solar radiation estimates and was only marginally important. This result concurs with the fact that the recovery of missing values of a weather variable was based mostly on the values of the weather variables the day before and the day after the day of interest. Use of the indicator variable showing presence or absence of rain is justified only if the user can determine whether the day with missing datum was rainy.

Both the coefficients and the structure of GMDH models may be valid only locally. Variables that are most important in estimating missing data may be the same for regions with similar climate, however. We plan to research this question in the near future. Although GMDH models gave much better estimates than climatology did in this study, the comparison with persistence estimates did not show decisive superiority of GMDH for all weather variables. For example, wind-run estimates from GMDH and from persistence had the same accuracy. GMDH should be tested in parallel with other forecast techniques when the estimation of missing data is needed.

In general, our results show that GMDH can be a useful tool in filling gaps in weather data from weather stations installed in the field. Selection of the predictor variables needs to be researched. The successful application of GMDH shows that historical weather data for a particular farm can be used to recover missing weather data important for on-farm decision support systems, including agricultural models.

REFERENCES

- Abdel-Aal, R. E., and M. A. Elhadidy, 1994: A machine-learning approach to modeling and forecasting the minimum temperature at Dhahran, Saudi Arabia. *Energy Int. J.*, **19**, 739–449.

- , and —, 1995: Modeling and forecasting the daily maximum temperature using abductive machine learning. *Wea. Forecasting*, **10**, 310–325.
- AbTech Corp., 1992–96: ModelQuest Version 4.0. User's Manual. [Available from AbTech Corporation, 1575 State Farm Blvd., Charlottesville, VA 22911.]
- Ashraf, M., J. C. Loftis, and K. G. Hubbard, 1997: Application of geostatistics to evaluate partial weather station networks. *Agric. For. Meteor.*, **84**, 255–264.
- Barron, A. R., 1984: Predicted square error—A criterion for the automated model selection. *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, S. J. Farrow, Ed., Marcel Dekker, 87–103.
- Blackie, J. R., and T. K. M. Simpson, 1993: Climatic variability within the Balquhidder catchments and its effect on Penman potential evaporation. *J. Hydrol.*, **145**, 371–387.
- Dinelli, D., 1995: What weather stations can do. *Landscape Manage.*, **34** (3), 6G.
- Dumais, R. E., Jr., and K. C. Young, 1995: Using a self-learning algorithm for single-station quantitative precipitation forecasting in Germany. *Wea. Forecasting*, **10**, 105–113.
- Elwell, D. L., and J. C. Klink, 1993: Ongoing experience with Ohio's Automatic Weather Station Network. *Appl. Eng. Agric.*, **9**, 437–441.
- Farrow, S. J., 1984: The GMDH algorithm. *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, S. J. Farrow, Ed., Marcel Dekker, 1–24.
- Fujioka, F. M., 1995: High resolution fire weather models. *Fire Manage. Notes*, **57**, 22–25.
- Lebow, W. M., R. K. Mehra, H. Rice, and P. M. Tolgalagi, 1984: Forecasting applications in agricultural and meteorological time series. *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, S. J. Farrow, Ed., Marcel Dekker, 121–147.
- Lin, Z.-S., J. Liu, and X.-D. He, 1994: The self-organizing methods of long term forecasting. I. GMDH and GMPSC methods. *Meteor. Atmos. Phys.*, **53**, 155–160.
- Mott, P., T. W. Sammis, and G. M. Southward, 1994: Climate data estimation using climate information from surrounding climate stations. *Appl. Eng. Agric.*, **10**, 41–44.
- Parker, D. D., and D. Zilberman, 1994: The adoption and use of information services: The case of CIMIS. Working Paper 736, 62 pp. [Available from University of California, Dept. of Agriculture and Resource Economics, Berkeley, CA 94720.]
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in FORTRAN: The Art of Scientific Computing*. 2d ed. Cambridge University Press, 992 pp.
- Reddy, V. R., B. Acock, and F. D. Whisler, 1995: Crop management and input optimization with GLYCIM: Differing cultivars. *Comput. Electron. Agric.*, **13**, 37–50.
- Setoodehnia, A., J. Y. Cheung, and H. Li, 1993: Hypercube clustering method: System modeling via modified group method of data handling. *Proc. 36th Midwest Symp. on Circuits and Systems*, Vol. 1, Detroit, MI, Institute of Electrical and Electronics Engineers, 422–425.
- Snyder, R. L., P. W. Brown, K. G. Hubbard, and S. J. Meyer, 1996: A guide to automated weather station networks in North America. *Adv. Bioclimatol.*, **4**, 1–61.