YRB 82-12

A COMPARISON OF MEASURED AND ESTIMATED METEOROLOGICAL DATA FOR USE IN CROP GROWTH MODELING*

Charles R. Perry, Jr.,
U. S. Department of Agriculture, Statistical Reporting Service

Julian L. Rogers,
U. S. Department of Agriculture, Federal Crop Insurance Corporation

ABSTRACT

Most crop growth simulation models and crop condition assessment models use, as part of their input, daily measurements for maximum temperature, minimum temperature, total precipitation, and solar radiation. Estimates of these variables are prepared for agriculture use from World Meteorological Organization surface reports and enhanced by polar orbiting satellites. If sufficiently accurate, use of this data may reduce the cost of obtaining reliable estimates of world crop conditions. These estimates are compared with actual daily meteorological data collected at various agricultural research facilities across the United States.

*Contribution of the Yield Model Development (YMD) project within the Agricultural Resources Inventory Surveys Through Aerospace Remote Sensing (AgRISTARS) program, a joint program of USDA, USDC, NASA, USDI, and AID.

INTRODUCTION

This paper outlines a study being conducted by the Yield Research Branch of the Statistical Reporting Service as part of the Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing (AgRISTARS) program, a joint USDA, NASA, NOAA, USDC, and AID research effort to determine the feasibility of integrating aerospace remote sensing technology into existing or future USDA data acquisition systems; some preliminary results are included to illustrate the techniques being used. The purpose of this study is to compare the U.S. Air Force Agromet data to actual measured daily meteorological data collected at various agricultural research facilities across the United States. The daily data elements being evaluated are maximum temperature, minimum temperature, total precipitation, and solar radiation. Measured data is not readily available for evaluation of potential evapotranspiration, and crop growth simulation. If the Agromet data--which is routinely available and prepared for agricultural use from World Meteorological Organization (WMO) surface reports and enhanced by polar orbiiting satellites--is accurate enough for plant and soil water modeling, data collection costs may be significantly reduced.

BACKGROUND

Interest in worldwide crop production and economic conditions has grown in recent years. In response, the U.S. Air Force has developed a complex model at Global Weather Central (GWC), Offutt Air Force Base, Nebraska, to provide specifically tailored daily meteorological data for agricultural use. Data to provide Agromet information comes from the WMO network surface reports and polar orbiting satellites measuring reflectance, radiance, and temperature. An automated cloud analysis model (3DNEPH) estimates the effect of clouds on the radiation balance.

Through an agreement between the Air Force (USAF) and the Department of Agriculture (USDA), this Agromet data is available for crop condition assessment and research use on a real time basis. An assessment of the accuracy of this data over a season has not been done.

DESCRIPTION OF DATA

The data elements described include only those for which an evaluation is being done. Agromet uses the 1/8 mesh AF GWC grid on a polar stereographic projection, which gives approximately 25 nm grid point spacing at 60° N, and all data elements are provided on this grid point basis. These estimates are currently provided for most of the United States and many areas of the Northern Hemisphere. Daily data for any land area of the world, Northern or Southern Hemisphere, could be provided by request and amendment to the existing USAF/USDA agreement. No historical data base of this data exists; data for the United States was not prepared prior to June 1981.

Maximum-Minimum Temperatures

Maximum and minimum temperatures are estimated every three hours from satellite temperature estimates. The highest and lowest estimates are saved for use at the end of the day. Surface reports (three hourly temperatures and daiy reported maximum-minimums) are used in an analysis with the satellite temperature estimates to create maximum-minimum temperatures for each grid point in each geographic region.

Precipitation

Precipitation reports from surface stations are assigned to the nearest grid point. Reported values are accumulated along with estimated amounts based on weather conditions from surface reports. Daily accumulations are made of the greater of the above amounts. Checks are made for convective precipitation and one-half of the daily accumulation is spread to adjacent grid points if no other precipitation is reported for those grids. Spreading is based on estimated amounts from the 3DNEPH cloud analysis. A ratio of reported/estimated amount is calculated for each grid point and these ratio values are spread using a linear distance weighting.

Reported precipitation is used for available points. At other points the R/E is used to determine a value. In a very few cases the quantative precipitation forecast (QPF) from the 3DNEPH is used. This is generally less than 12 grid points for an area.

Solar Radiation

Clear sky direct solar radiation is calculated from long standing, well-known equations. The clear sky solar radiation is adjusted for cloudiness using 15 cloud layers in the 3DNEPH cloud model. A detailed explanation of Agromet net solar radiation computation is given in USDA ETAC/TN-81-001, March 1981.

Other Agromet Data

Other data elements produced by Agromet are not being evaluated at this time. A description of these data can be obtained from ETAC/TN-81-001.

DATA COLLECTION

The Agromet data being used in this study was processed in the Foreign Agricultural Service (FAS), Foreign Crop Condition Assessment Division (FCCAD), Houston, Texas. Measured data was collected by various researchers around the country and assembled by Dr. J. Ritchie, USDA/SRS, Blackland Research Center, Temple, Texas, for evaluation. The data associated with the Blackland Research Center are displayed graphically in Appendix A, along with several graphs that give a "feel" for the character of these data.

Air Force Agromet Data

Agromet is processed continuously (every three hours) at GWC, Offutt Air Force Base, Nebraska. Daily at 2400 GMT the daily Agromet Data Summary is prepared. These daily data are assembled every Monday into a weekly (Monday through Sunday) data set and a magnetic tape is Air Expressed to FCCAD, Houston, Texas. The data is processed each Tuesday and is available for operational use generally within 60 hours after preparation at GWC. Data for comparison to measured data has been extracted from the FCCAD disk file for use in statistical analysis.

Measured Data

Daily maximum-minimum temperatures, total preciitation, and net solar radition is routinely collected by various researchers around the country. Through private communication these data were sent to Dr. J. Ritchie, USDA/ARS, Temple, Texas, monthly for assembly and conversion to computer compatible media. These data were furnished to the Yield Model Development (YMD) project, Houston, for use in statistical evaluation of the Agromet data. YMD use of this data is limited to statistical analysis.

DATA EVALUATION

Each of the four Agromet data elements (maximum temperature, minimum temperature, precipitation, and solar radiation) is being compared to the measured data elements for each day.

Grid Cell Data

Data elements for each grid cell in which a research location is situated are being compared against the measured data element for the same day. These comparisons are being done on a monthly basis as well as for the overall annual period (June 1981-May 1982). Some averages or accumulations of some data elements over a period greater than one day will be used for comparisons.

Interpolated Data

Data elements for the research location grid cell and each grid cell that is immediately adjacent to that grid cell (nine cells) will be interpolated by a weighted linear distance method to obtain a value for the measuring location for comparison to each measured data element. Statistical analysis will then be performed as described for the grid cell data.

STATISTICAL ANALYSIS OUTLINED

Analytic and graphical methods are being used to characterize the error structure of the Agromet estimators. The results will be presented numerically and graphically.

The mean, variance, and other standard statistics for each set of differences described above are being computed. The mean and variance of the differences provide estimates of the bias and variance expected in using the Agromet data, assuming that the difference estimates the error made by using the Agromet data instead of the "true" data element. Test for month and station effects will be examined using a two factor linear model with month and station random factors. Similar non-parametric test will be performed and the results compared with results from the parametric tests.

Both historgrams and time plots are being made and visually examined for each set of differences. These graphs generally provide additional insight into the expected behavior of the estimators, and they may indicate that additional statistical testing and characterization of the estimation errors to be advantageous.

Since the measured data provides precise measurements of the data values at a specific point; and the Agromet data provides estimates of data values over a large area (no smaller than 25x25 nm), differences in the two data sets are to be expected. The purpose of this analysis is to estimate these differences and to understand the behaviour of these differences. Such an understanding is necessary before using the Agromet data for large area agricultural estimation without regard or concern as to whether or not a ground station is available over the area to provide meteorological measurements.

SOME PRELIMINARY RESULTS

One way to characterize the errors made in using the daily Agromet cell estimates as an estimator of the station point measurements is to fit a general linear model to the daily differences between these measurements. Since the station locations are a subset (more or less randomly located in the major agricultural land of the U.S.) of a much larger set of points for which we wish to make inferences (develop crop growth and other models for), it is clear that the classification variable station location is to be considered a random effect. The situation with the classification variable month is not so obvious.

The months themselves are not a random sample of all months; they are the twelve consecutive months for which the data was gathered. However, the effect on the differences may be considered a random effect because we are not interested in comparing specific months; we are only interested in ascertaining how much of the estimation error can be attributed to month to month changes in differences. The analysis for the variable daily maximum temperature are presented as an example of the technique used and the results obtained.

The two factor random effects model was fit to the data. The components of variance associated with the residual error, the month effect, the station effect, and the station and month interaction were estimated by three standard methods, and hypothesis were tested using several analogous type sums of squares to ascertain if any of the components of variance could reasonably be considered to be zero. The results for testing the hypothesis that a variance component is zero are sumarized in Table 1; the results are given using type IV analogous sums of squares the results were identical when other type sums of square were used. Table 2 gives the estimates for the variance components for each of the three methods employed. The obvious conclusion is that the variance component associated with the classification variable should be considered zero. Following standard statistical procedure the classification variable month was dropped from the analysis and the reduced model refitted to the data, the hypothesis retested, and the variance components reestimated. The results are presented in Tables 3 and 4.

The obvious conclusion one must draws from the results given in Tables 1-4 is that the variance components associated with the station location and the station month interaction are not zero and that both contribute significantly to the total error made in using the Agromet daily cell maximum-temperature estimate as an estimator of the associated station point maximum-temperature measurement. It is equally obvious that there is no reason to suspect the variance component associated with the classification variable month contributes significantly to the overall error structure; hence, it is reasonable to consider it zero.

From Table 4 we observe that the residual variance, the station variance, and the station*month variance attribute respectively about 85%, 9%, and 7% to the total variance. It is perhaps instructive to observe that one can expect that about ninety percent of the time the daily cell grid maximum temperature estimate is within 9.1 degrees celsius (9.1=1.64 times estimated total variance) of the associated point station measurement for daily maximum temperature; the contribution of the residual error to this value is 8.4 degrees celsuis.

TABLE 1: ANALYSIS OF VARIANCE FOR THE FULL MODEL USING TYPE IV SUMS OF SQUARES TO TEST

H_0: VAR(MONTH) = 0; H_0: VAR(STATION) = 0; H_0: VAR(STATION*MONTH) = 0

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	COMPUTED F_0	PROB $F > F_0$	EXPECTED MEAN SQUARE
MONTH	11	704.5	0.67	0.76	$\sigma_E^2 + 29.15\,\sigma_{S*M}^2 + 297.97\,\sigma_M^2$
STATION	10	9712.5	10.18	0.0001	$\sigma_E^2 + 28.89\,\sigma_{S*M}^2 + 322.41\,\sigma_S^2$
STATION*MONTH	104	9921.4	3.67	0.0001	$\sigma_E^2 + 29.15\,\sigma_{S*M}^2$
ERROR	3551	92400.9			σ_E^2

VAR (ERROR) $\equiv \sigma_E^2$

VAR (STATION) $\equiv \sigma_S^2$

VAR (STATION*MONTH) $\equiv \sigma_{S*M}^2$

VAR (MONTH) $\equiv \sigma_M^2$

8/13/82

TABLE 2: VARIANCE COMPONENT ESTIMATES
FOR THE FULL MODEL

VARIANCE COMPONENT	VARIANCE COMPONENT ESTIMATION PROCEDURE		
	TYPEI SS	MIVQUEO	MAXIMUM LIKELIHOOD
$\widehat{VAR}$ (MONTH)	-0.12	-0.12	0.00
$\widehat{VAR}$ (STATION	2.68	2.69	2.43
$\widehat{VAR}$ (STATION*MONTH)	2.38	2.39	2.23
$\widehat{VAR}$ (ERROR)	26.02	26.00	26.02

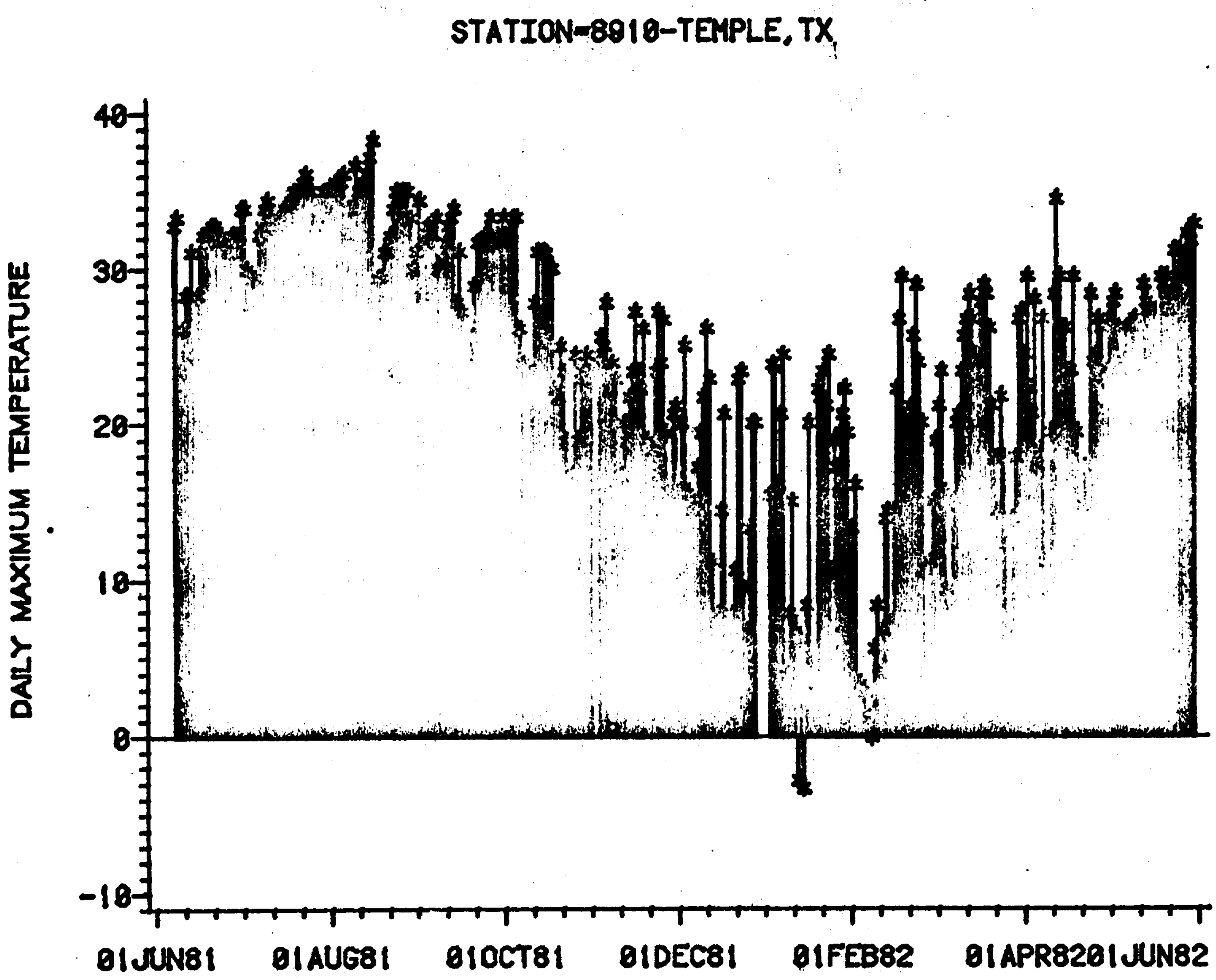
TABLE 3: ANALYSIS OF VARIANCE FOR THE REDUCE MODEL USING TYPE IV SUMS OF SQUARES TO TEST

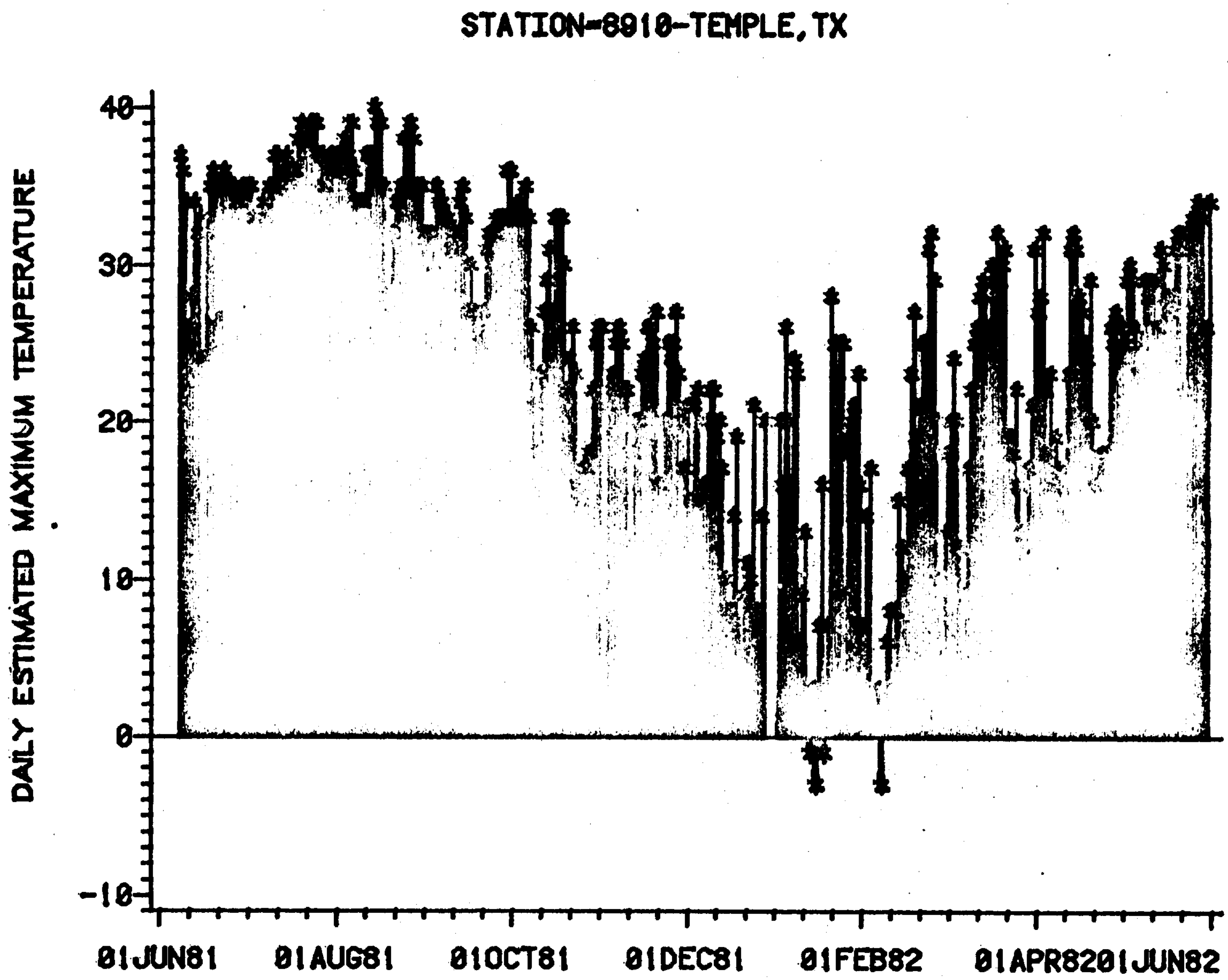
H_0: VAR(STATION) = 0 ; H_0: VAR(STATION*MONTH) = 0

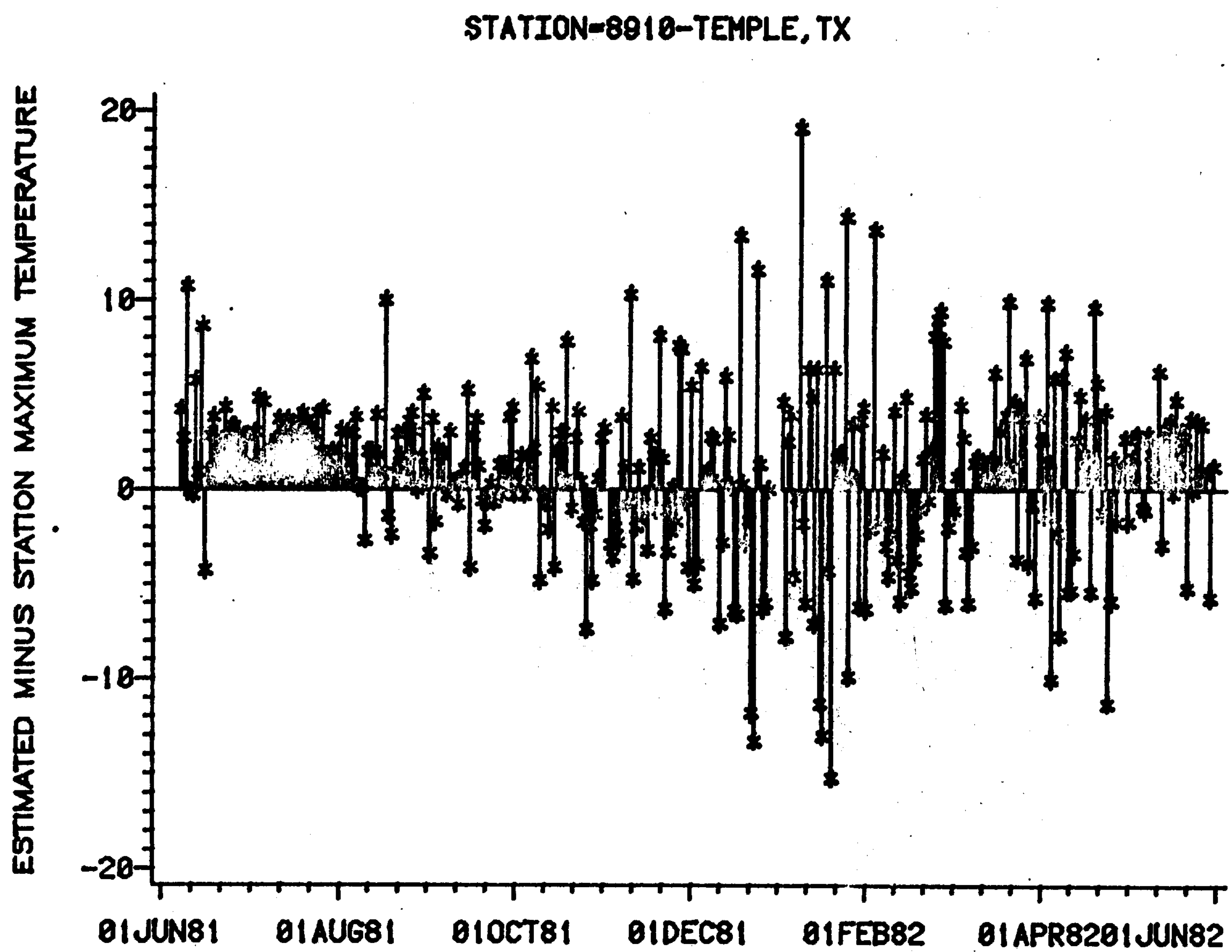
SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	COMPUTED F_0	PROB $F > F_0$	EXPECTED MEAN SQUARES
STATION	10	9897.71	10.83	0.0001	$\sigma_E^2 + 28.89\ \sigma_{S*M}^2 + 330.39\,\sigma_S^2$
STATION*MONTH	104	10509.79	3.51	0.0001	$\sigma_E^2 + 29.16\ \sigma_{S*M}^2$
ERROR	3551	92400.94			σ_E^2

TABLE 4: VARIANCE COMPONENT ESTIMATES
FOR THE REDUCED MODEL

VARIANCE COMPONENT	VARIANCE COMPONENT ESTIMATION PROCEDURE					
	TYPE I SS		MIVQUEO		MAXIMUM LIKELIHOOD	
	ESTIMATE	PERCENTAGE	ESTIMATE	PERCENTAGE	ESTIMATE	PERCENTAGE
$\widehat{VAR}$ (STATION	2.71	8.8	2.70	8.7	2.43	7.9
$\widehat{VAR}$ (STATION*MONTH)	2.24	7.2	2.27	7.3	2.23	7.3
$\widehat{VAR}$ (ERROR)	26.02	84.0	26.00	84.0	26.02	84.8
$\widehat{\text{TOTAL VARIANCE}}$	30.97	100.0	30.97	100.0	30.68	100.0

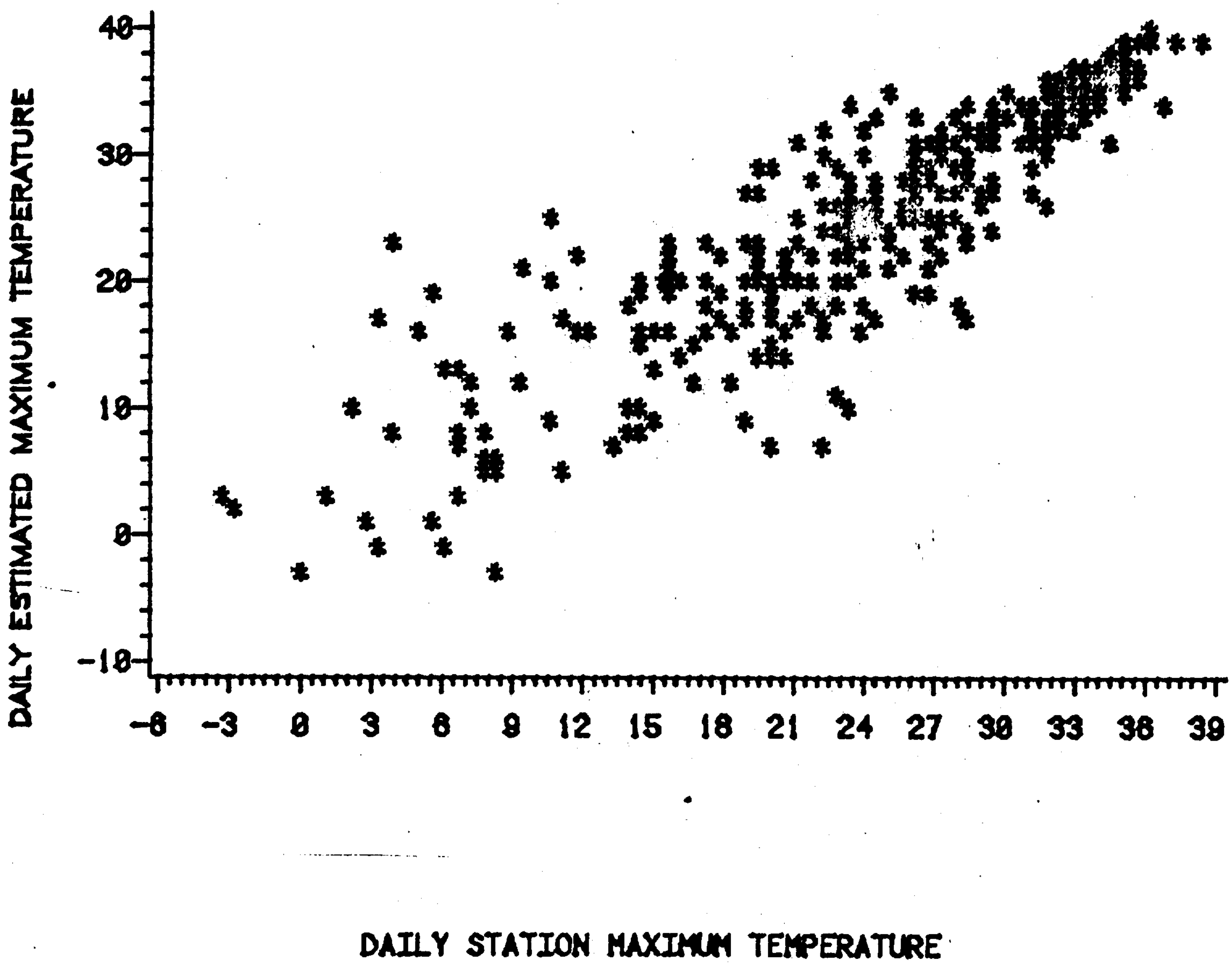

STATION-8910-TEMPLE, TX
DAILY MAXIMUM TEMPERATURE
40
30
20
10
0
-10
01JUN81
01AUG81
01OCT81
01DEC81
01FEB82
01APR82
01JUN82


STATION=8910-TEMPLE,TX
DAILY ESTIMATED MAXIMUM TEMPERATURE
40
30
20
10
0
-10
01JUN81
01AUG81
01OCT81
01DEC81
01FEB82
01APR82
01JUN82

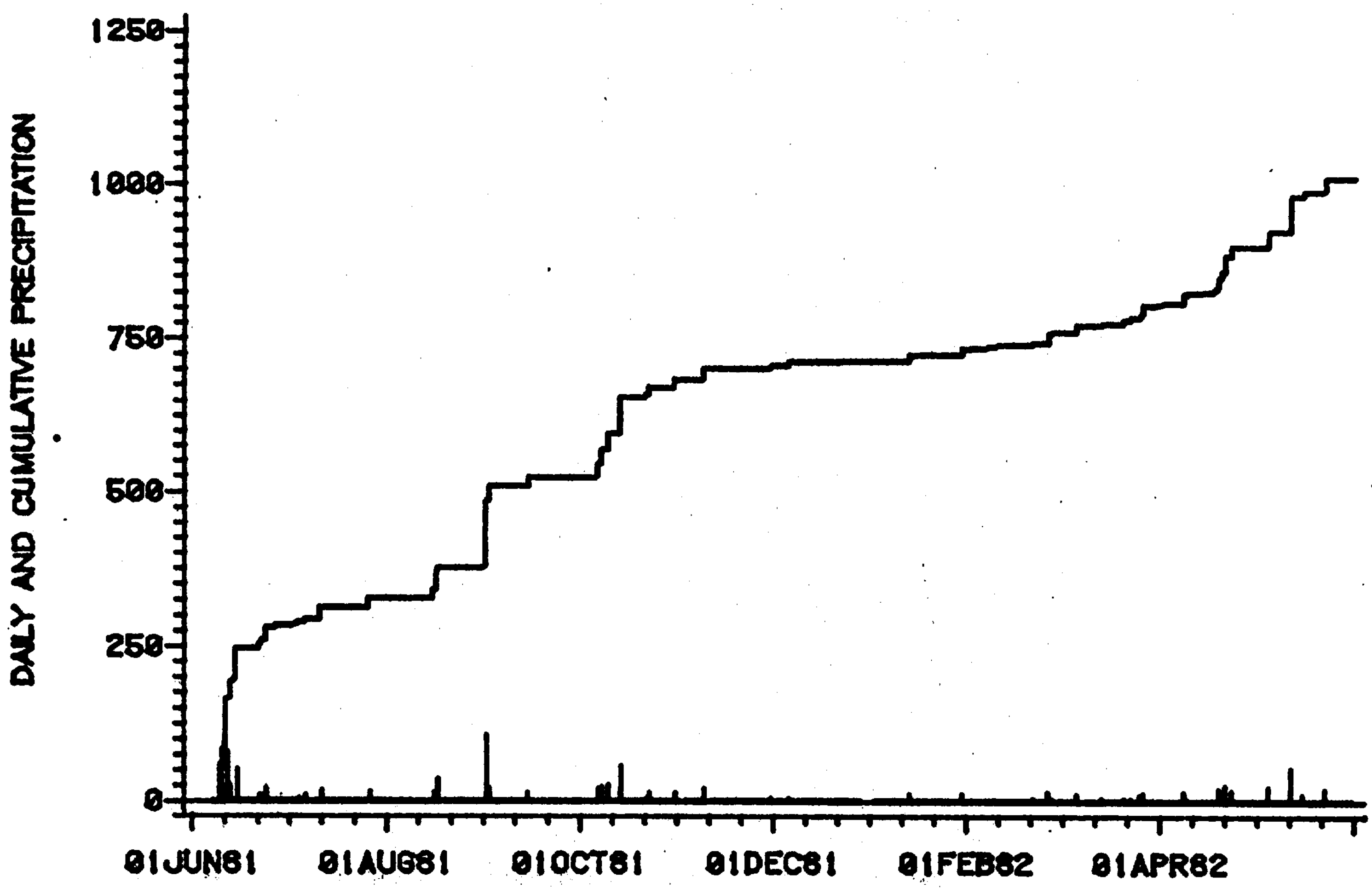


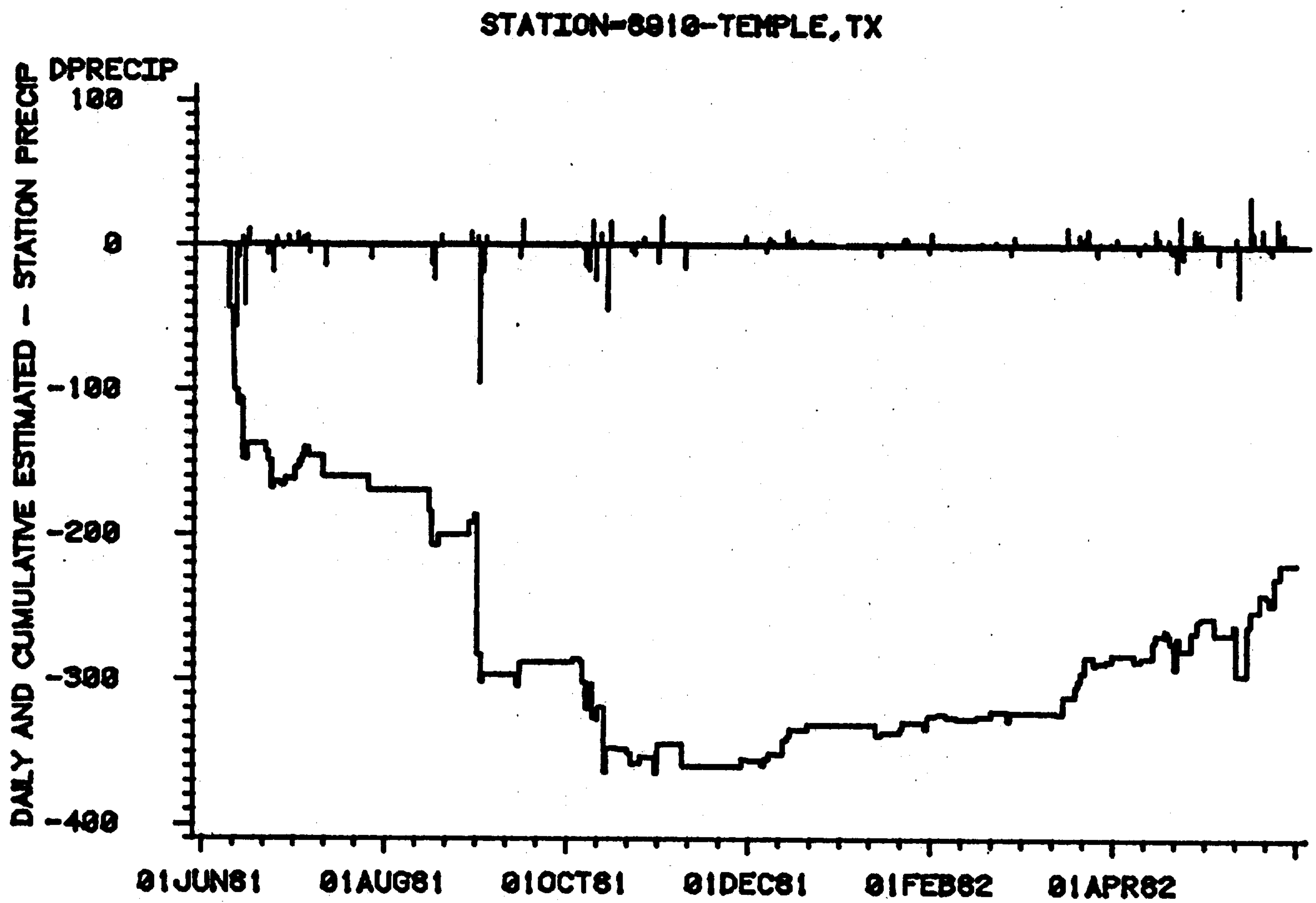
STATION=8910-TEMPLE,TX
ESTIMATED MINUS STATION MAXIMUM TEMPERATURE
20
10
0
-10
-20
01JUN81
01AUG81
01OCT81
01DEC81
01FEB82
01APR82
01JUN82

STATION-8910-TEMPLE,TX



STATION-8910-TEMPLE, TX




STATION-8910-TEMPLE, TX
DPRECIP
DAILY AND CUMULATIVE ESTIMATED - STATION PRECIP
100
0
-100
-200
-300
-400
01JUN81
01AUG81
01OCT81
01DEC81
01FEB82
01APR82

References

1. USAF ETAC/TN81/001 Soil Moisture and Agromet Models.

2. Gridded Meteorological Data Extraction System, Users Guide.

3. S. R. Searle (1971). Linear Models; Wiley, New York.

4. R. T. Freund and R. C. Littel. SAS for Linear Models. SAS Institute, Cary, North Carolina.

5. SAS Users Guide 1979 Edition. SAS Institute, Cary, North Carolina.