United States
Department of
Agriculture

National
Agricultural
Statistics
Service

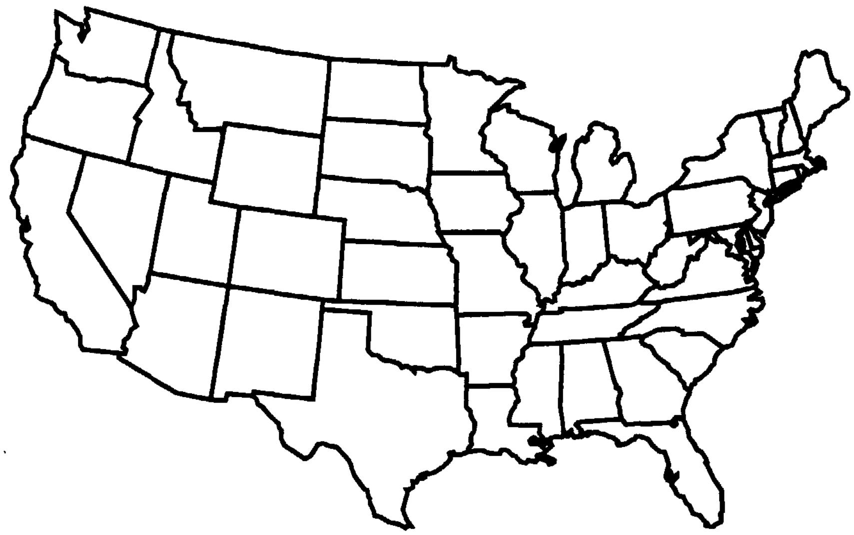
Research Division

SRB Research Report
Number SRB-94-03

February 1994

A Quality Assessment of Near Real-Time Precipitation Data Used to Make Crop Yield Forecasts

M. Denice McCormick Myers



A QUALITY ASSESSMENT OF NEAR REAL-TIME PRECIPITATION DATA USED TO MAKE CROP YIELD FORECASTS, by M. Denice McCormick Myers, Sampling and Estimation Research Section, Survey Research Branch, Research Division, National Agricultural Statistics Service, United States Department of Agriculture, Washington, DC 20250-2000, February 1994, Report No. SRB-94-03.

ABSTRACT

The National Agricultural Statistics Service (NASS) uses regression models to forecast yield for crops such as corn, soybeans and winter wheat. Analyses were conducted on the use of precipitation data in these regression models (McCormick and Birkett 1992, and McCormick 1993). Precipitation data are obtained from two sources. The National Climatic Data Center (NCDC) supplies historic precipitation data used for developing regression model parameters. The Climate Analysis Center (CAC) supplies current year near real-time precipitation data from first-order synoptic and airways stations that can be used as regression model input in a production setting. Two issues of data quality are addressed by this study. First, CAC weather station density is sparse across the U.S. in many major agricultural areas as compared to that of NCDC (McCormick, 1993). Second, CAC data are preliminary, semi-edited data. As a result, significant differences exist between NCDC (final edited data) and CAC (preliminary semi-edited data) on individual station and State aggregated term levels. Although relatively small, significant differences due to density were found for corn on the regional level, but larger differences were found for winter wheat. Significant differences due to editing are smaller and have no practical importance in crop yield forecasting. This paper evaluates these two sources of error in near real-time precipitation data to determine their effects on regional forecast accuracy. Results indicate that corn, soybean and winter wheat models based on CAC near real-time data are just as accurate as NCDC historic data based models. However, continued quality control efforts of the CAC near real-time data base and close supervision of crop yield models using these data are recommended.

KEY WORDS: **Precipitation data; regression models; forecast accuracy.**

> This paper was prepared for limited distribution to the research community outside the U.S. Department of Agriculture. The views expressed herein are not necessarily those of NASS or USDA.

ACKNOWLEDGEMENTS

The author wishes to thank James Laver, Chief, Analysis and Information Branch, and David Miskus, Climate Analysis Center; Grant Goodge, Head, Quality Control, National Climatic Data Center; Raymond Motha, Supervisory Meteorologist, Robert Stefanski and Mark Brusberg, World Agricultural Outlook Board; William Iwig, Head, Sampling and Estimation Research Section, and Phil Kott, Chief Research Statistician, Research Division, National Agricultural Statistics Service for all their support, assistance and commitment to this research project.

TABLE OF CONTENTS

SUMMARY

The benefits of using Climate Analysis Center (CAC) data to derive terms used in crop yield forecasting are clear because of its near real-time availability and low cost. However, significant differences exist between terms derived from CAC versus National Climatic Data Center (NCDC) historic data. Part of the differences occurs because NCDC has a much denser network of weather stations (McCormick, 1993). In comparison, this network contains approximately 5,000-6,000 stations compared to the CAC network of about 300-400 stations, making it more than ten times larger. Another measurable difference occurs because CAC data are reported quickly, passing through limited editing and quality control checks. NCDC data undergo robust quality control and editing procedures prior to publishing for public use. The National Agricultural Statistics Service (NASS), quite naturally, has a valid concern that either of these aspects could create problems in crop yield forecast accuracy. In order to examine both sources of potential error, precipitation terms were derived from each data source to develop model parameters for making corn, soybean and winter wheat yield predictions. Results regarding the first issue of differences in station density showed that:

(1) CAC regional precipitation values are significantly different than NCDC values for corn and winter wheat at the $\alpha = 0.10$ level of significance. Precipitation mean differences in the corn region for June and July are 0.09 inches for both months. The May forecast period winter wheat mean difference is larger, at 0.39 inches. Precipitation mean differences in the soybean region are 0.09 and 0.02 inches for the August and September forecasts, respectively. Neither is significantly different. Overall, the corn and soybean differences are relatively small and, if significantly different, have no practical importance since those differences did not affect forecast accuracy. The winter wheat difference indicates the CAC data should be used with caution for that crop.

(2) Regional crop yield models developed using CAC-based terms were just as accurate as those based on NCDC terms. CAC models capture the impact of precipitation on final yield comparably to NCDC models even though CAC values are based on only about one-tenth the number of stations as NCDC values and only provide coverage on approximately 70% of crop harvested acres.

NASS has been collecting data from the CAC Climate Dial-Up Service (CDUS) monthly since March 1992. Our historic NCDC data base that contained NCDC data from 1980 through 1991 was updated with 1992 NCDC-edited published data that replaced 1992 CAC semi-edited preliminary data. To analyze the second issue of terms having significant differences due to editing changes, a comparison was made between data sources of 1992 monthly terms that are used in the early season, May 1 (winter wheat) and August 1 (corn and soybean) crop yield forecast models. Keep in mind that this portion of the analysis is only based on one year (1992) of data. Therefore, making conclusions is not possible. However, the level of differences is documented. Differences in precipitation terms due to

editing have minimal effects on regional level forecasts, based on this single year of analysis.

Data quality is extremely important. After examining the combined issues of station density and differences due to data editing, four recommendations are offered to ensure successful data usage:

(1) Use CAC data as current year input for corn and soybean forecasts, but with caution for winter wheat forecasts,

(2) update the NCDC historic precipitation data base annually until a sizeable CAC data series can be fully evaluated,

(3) continue to maintain the CAC monthly data series,

(4) and within the bounds of inter-Agency cooperative efforts, work with the CAC to improve the quality of near real-time data to minimize changes due to editing through further evaluation and monitoring of data editing procedures each year.

A QUALITY ASSESSMENT OF NEAR REAL-TIME PRECIPITATION DATA USED TO MAKE CROP YIELD FORECASTS

M. Denice McCormick Myers

INTRODUCTION

The National Agricultural Statistics Service (NASS) evaluated the addition of precipitation terms to regression models used to forecast crop yields at the regional level (McCormick and Birkett, 1992, and McCormick, 1993). Previously, these models were based solely on survey data-related terms. The survey data were collected from randomly selected sample plots in randomly selected fields in States within each crop region. For fall harvested crops, the surveys are conducted monthly and forecasting starts August 1. For winter wheat, the initial monthly forecast is made May 1. Linear or quadratic regression models are developed for each month and crop, based on historic data relating a regional survey variable to final yield. Data from the current survey are then aggregated to the regional level and used in the model to forecast final yield.

NASS used two different sources of precipitation data for this research effort. The Climate Analysis Center (CAC) of the National Weather Service provides daily data in a near real-time mode for the current and past three years from a fairly sparse network of about eight to twelve automated first-order "Synoptic" (primary) and "Airways" (secondary) stations per State. The National Climatic Data Center (NCDC) of the National Weather Service provides historic data from a much denser network of about 80 to 120 stations per state, about ten times the number that CAC provides, with almost one station in every county. The NCDC network includes the CAC automated station network plus a large number of other manually operated stations. Manual operators are often volunteer local weather reporters such as local farmers, universities and agricultural extension agents. These additional stations are commonly called the "Cooperative" station network. NCDC uses robust quality control measures to edit the data. Typical quality control measures involve checking for decimal point location in data reports. For example, a reporter could report 25 inches when he intended to report 2.50 inches of precipitation. Comparisons of stations located in surrounding areas called "Buddy Checks" highlight possible data anomalies. Additionally, missing data are checked and late reports are incorporated into the data base. Because of the intensive effort to compile this data set, there is a minimum four-month lag until it becomes available making its usage impractical for current year crop yield forecasting. Therefore, in previous research and applications done by NASS, historic NCDC data were used to develop the regression model parameters and CAC data were used as input to the model for the current year forecast. NCDC data for the current year were obtained when available to update the data base and to develop new models for the next crop year.

The sparse density of CAC stations and possible errors due to a lack of intense editing associated with the near real-time CAC data are concerns to NASS regarding its use in crop yield forecasting. However, easy access and near real-time availability of the CAC data are attractive. The sparse CAC network within States still provides 85 stations for current year corn forecasting, 71 stations for soybean forecasting, and 179 stations for winter wheat forecasting at the regional (multi-State) level. The average number of stations per State is approximately nine for corn and soybean States and eleven for winter wheat States. Most States have at least one CAC station per Agricultural Statistics District (ASD), formerly called "Crop Reporting District." An ASD is a region consisting of contiguous counties within each State that is characterized by similar agricultural characteristics, such as homogeneous soil, cropping practices and climatic patterns (Arends, W., et al. 1983).

Fast reporting time is the primary advantage of using CAC data. The CAC has the capacity to collect data on an hourly basis and calculate a daily summary by the following morning for access by the Joint Agricultural Weather Facility (JAWF) through its agreement with the World Outlook Board, an agency of the U.S. Department of Agriculture (Motha and Heddinghaus, 1986). For example, data that was collected on January 13, was ready by January 14. There are trade-offs between speed and accuracy. However, it would be incorrect to report that the CAC does not concern itself with accuracy. There are automated quality control efforts that are done during daily data processing and summarizing, some of which have lengthy descriptions. A simple description of one of them involves the maintenance of a library listing of stations that is updated bi-annually by "flagging" inoperative and "suspicious" stations. If time and resources permit, some of the other measures taken include manual data examination on a station-by-station level basis, data plots, Buddy Checks, comparisons to River Forecast Center (RFC) data and comparison to historic data by percent of normals. Operationally, preliminary daily data reports are summarized for the "Monthly Summary" that is updated on the Climate Dial-Up Service (CDUS) on the first day of each month and released on the second for public use. Corrections are made a few days later from data received and comparisons made to "CLIMAT" data (monthly totals reported by stations used to check daily data reports) between the third to sixth day of each month. Because of time limitations before regional crop yield forecasts, the first reports are used to approximate the current year data set.

This study examines the differences between the NCDC and near real-time CAC precipitation values at regional levels. If large differences occur, the use of a CAC value as input to an NCDC-based model is not appropriate. Differences between regional precipitation values from NCDC and CAC sources are due to density of the weather stations and editing changes. Density differences and their impact are studied by using 1980 to 1991 NCDC data and a subset of the NCDC data corresponding to the density of the CAC stations. Editing differences are examined by comparing 1992 near real-time CAC data to matching NCDC data at individual station and regional levels. This was the first year NASS accessed near real-time CAC data, so only one year of data had been collected. Consequently, no conclusions can be made about the expected difference or effect over years. The level of the 1992 differences is documented.

Analysis for this study was conducted at the regional level over the ten Corn Objective Yield (OY) States Illinois (IL), Indiana (IN), Iowa (IA), Michigan (MI), Minnesota (MN), Missouri (MO), Nebraska (NE), Ohio (OH), South Dakota (SD), and Wisconsin (WI), the eight former Soybean OY States Arkansas (AR), Illinois (IL), Indiana (IN), Iowa (IA), Minnesota (MN), Missouri (MO), Nebraska (NE), and Ohio (OH), and the fifteen former Winter Wheat OY States Arkansas (AR), Colorado (CO), Idaho (ID), Illinois (IL), Indiana (IN), Kansas (KS), Missouri (MO), Montana (MT), Nebraska (NE), Ohio (OH), Oklahoma (OK), Oregon (OR), South Dakota (SD), Texas (TX), and Washington (WA). The word "former" is used because this reflects the set of States that made up the crop regions in the 1991 crop year.

Specifically, the objectives of the study are the following:

(1) Determine if the differences in the densities of the NCDC and CAC networks cause differences in the regional level precipitation values that would prohibit the use of a CAC value as input to a NCDC-based model.

(2) Document the level of differences due to editing the 1992 near real-time CAC data.

DATA

Current year precipitation data are obtained directly from CDUS at the CAC each month over a modem to a personal computer. CAC is the best known source of near real-time data. However, as noted previously, data are characterized by few edit or quality control checks for the preliminary report. Weather stations that supply the data are sparsely located across the country with a tendency to be located at airports and Army or Air Force airfields, and in urban rather than agricultural areas. The network is capable of providing daily, weekly, or monthly accumulated data. NASS uses CAC's monthly summarized data for each weather station. It then aggregates the data to monthly totals on ASD, State, and regional levels. Table 1 provides a summary of CAC station coverage by crop and State based on 1991 harvested acreage. The coverage statistic C_s is the percentage of harvested acres by crop within a State that is represented by a CAC weather station out of the total State harvested acres:

$$C_s = \frac{\sum_{d=1}^{D_s} I_{sd}(A_{sd})}{\sum_{d=1}^{D_s} A_{sd}} 100$$

where

C_s = The State coverage percentage for State s,

D_s = number of districts in State s,

I_{sd} = 1, for State s, district d if the district is covered by an

Table 1: CAC Coverage

State	RANK 91[†]	#Stations	C_s[‡]	100-C_s
CORN:				
IL	2	10	64	36
IN	5	6	69	31
IA	1	8	87	13
MI		17	91	9
MN	4	7	54	46
MO		10	62	38
NE	3	9	79	21
OH		13	74	26
SD		8	65	35
WI		7	65	35
REGION TOTAL:		**85**	**72**	**28**
SOYBEANS:				
AR		8	54	46
IL	1	10	54	46
IN	4	6	70	30
IA	2	8	86	14
MN	3	7	46	54
MO		10	59	41
NE		9	74	26
OH	5	13	77	23
REGION TOTAL:		**71**	**65**	**35**
WINTER WHEAT:				
AR		8	53	47
CO	4	10	100	0
ID		5	73	27
IL		10	38	62
IN		6	77	23
KS	1	12	86	14
MO		10	58	42
MT		11	100	0
NE	5	9	72	28
OH		13	76	24
OK	2	9	83	17
OR		12	54	46
SD		8	66	34
TX	3	41	75	25
WA		15	100	0
REGION TOTAL:		**179**	**79**	**21**

[†] Rank 91 = rank of 5 largest states based on 1991 harvested acres.
[‡] C_s = State weather station coverage percentage as defined in equation on page 8.

		automated station,
		0, otherwise, and
A_{sd}	=	Total harvested acres, for state s, district d.

Harvested acres of a district are considered covered if the district contains a CAC station. The under-coverage statistic is then (100 - C_s). Regional coverage statistics C are also provided:

$$C = \frac{\sum_{s=1}^{S}\sum_{d=1}^{D_s} I_{sd}(A_{sd})}{\sum_{s=1}^{S}\sum_{d=1}^{D_s} A_{sd}} 100$$

where

S	=	number of states in region (for the crop).

Table 1 lists each CAC State and Regional Coverage statistic.

Table 2: Precipitation Term Month

Crop	Forecast Date	Precipitation Term
Corn	August 1	July
	September 1	June
Soybeans	August 1	July
	September 1	August
Winter Wheat	May 1	August (lag 1 year)+ March (current)

The precipitation data are summarized to regional totals for particular months or groups of months (McCormick 1993). Different months are used for different crops and monthly forecasts as documented in Table 2. The survey data collected are also aggregated to the regional level. The procedures used to aggregate both the precipitation and survey data are documented in Appendix A.

METHODOLOGY

Evaluating Differences in Regional Precipitation Due to Density

To evaluate differences in the NCDC and CAC regional precipitation values due to differences in network densities, a subset of stations in the full NCDC data set that matched the location of CAC stations was created. A proxy CAC data set from 1980 through 1991 was built from the NCDC data set by extracting all stations located in cities appearing in the February 1993 CAC network. February 1993 was the most current data set collected at the time this study began. For the purposes of this analysis, an assumption was made that the CAC network always consisted of the same station locations historically. Usually this is the case, however in some instances stations are moved to areas having similar climatic and geological characteristics. During this part of the analysis city (location) was the only means of "matching" stations between the data sets at that time. Monthly reports of precipitation for matching cities in the CAC proxy data set that had multiple stations were averaged together. For example, Kansas City, Missouri had one station in the February 1993, CAC data set, but two Kansas City stations appeared in the NCDC data set from 1980 through 1991. The reports were averaged for that location each year. Later, station matching will be done by NCDC station reference number for the second part of this analysis (See "Evaluating Differences in Regional Precipitation due to Editing.")

Regional level precipitation values were created for each network. Precipitation values were compared using a paired t-test. Kaiser and Sebaugh (1984) used this approach previously to compare near real-time and historic weather data to forecast barley and spring wheat yields in North Dakota from 1970 to 1979. Historic parameter values were derived from 1970 to 1978 NCDC data. Current year input values were derived from the 1979 CAC data from stations in North Dakota. This was the first year that data from the CAC was available since it was founded in 1979. At that time, the study concluded that values derived from near real-time data were not equivalent to historic values.

Further, the report noted that the small number of stations in operation in North Dakota prohibited them from using CAC data for near real-time crop yield forecasting.

In additional analysis, regression model parameters were built based on both data sets for each year in the data series. For crop regions where the precipitation values are comparable, the performances of the regression models are expected to be comparable. In the absence of measurement error of the independent variables, differences due to density can be evaluated using multiple linear regression models of the form:

1: $Y_t = \beta_o + \beta_1 Z_t + \epsilon_t$

2: $Y_t = \beta_o + \beta_1 Z_t + \beta_2 Z_t^2 + \epsilon_t$

3: $Y_t = \beta_o + \beta_1 Z_t + \beta_2 P_t + \epsilon_t$

4: $Y_t = \beta_o + \beta_1 Z_t + \beta_2 Z_t^2 + \beta_3 P_t + \epsilon_t$

where:

Z_t = the regional aggregated survey variable as documented in Appendix A, and

P_t = the regional aggregated precipitation variable also as documented in Appendix A.

Model 2 is the official model used by NASS to forecast August corn and soybeans and September soybeans. However, Model 1 is the official model used to forecast September corn and May winter wheat.

Models 3 and 4 each use one monthly precipitation term. Analysis was conducted previously (McCormick 1993) to determine which month or combination of months from the growing season provided optimal forecasting capability. Also, models with multiple monthly precipitation terms were examined.

Model Evaluation Criteria

The primary model evaluation criterium is the set of prediction intervals (PI) that correspond to the years with the minimum, median, and maximum yields over the twelve years in the study for each crop. These years for corn were 1983, 1989, and 1986, respectively; for soybeans 1988, 1981, and 1990, respectively; and for winter wheat were 1989, 1985, and 1983 respectively. A second criterium is the adjusted coefficient of determination, R_a^2, which provides a measure of correspondence between predicted and actual yields. Both the PI and R_a^2 are based on the sum of squared differences from the least squares analysis used to derive the model parameters.

1. The prediction interval (PI) refers to half the confidence interval length for the predicted value of a future Y for a given future year o.

That is, at the α significance level:

$$PI = t(1-\frac{\alpha}{2};n-1-p)SD(\hat{Y}_o),$$

where

$$SD(\hat{Y}_o)=s[(x_o'(X_o'X_o)^{-1}x_o) + 1]^{\frac{1}{2}},$$

s = $(\text{residual MSE})^{1/2}$,
x_o = relevant p-dimensional row vector of independent variables for year o (for example, in Model 3: p=3, $x_o = [1, Z_o, P_o]$),
X_o = relevant (n-1 X p) matrix of independent variables (excludes x_o),
n = number of years, and
p = number of parameters.

The X_o matrix excludes the row vector x_o, so that the PI reflects the accuracy expected in an operational model where current year data are not included in the model development. A significance level of 0.32 provided t values near 1.0. Consequently, the future Y will fall within the calculated PI of the predicted Y approximately 68% of the time.

2. R_a^2 is used as a goodness-of-fit test for each model with an adjustment made for the corresponding degrees of freedom (Draper and Smith 1981).

R_a^2 is calculated as:

$$R_a^2 = 1-\frac{(RSS_p)/(n - p)}{(CTSS)/(n - 1)},$$

where

RSS_p = the residual sum of squares taking the changing number of parameters into account,
CTSS = the corrected total sum of squares,
n = the number of years, and
p = the number of parameters.

Outlier Identification

Since the purpose of the models is to make forecasts, the rstudent statistic (also called the studentized residual) was used to help identify outliers to be excluded from the model. This statistic, recommended in Belsley, Kuh and Welsh (1980), is similar to the standardized residual:

$$r_{si} = \frac{r_i}{s\sqrt{1-h_i}},$$

where

r_i = i^{th} residual,
s = $(\text{residual MSE})^{1/2}$, and
h_i = $x_i'(X'X)^{-1}x_i$

S is replaced by s(i), where S(i) is the estimate of σ with the i^{th} observation deleted. In a forecasting model, the rstudent statistic measures how the distance of the forecast from the observed Y in terms of prediction standard errors. Observations with absolute values of rstudent greater than 3.0 were identified as outliers. The rstudent statistic is distributed closely to the t-distribution with n-p-1 degrees of freedom. For corn, twelve years of data (1980-1991) minus 1988 outlier year) were used, leaving eight degrees of freedom for the analysis for each forecast period. The soybean analysis had nine degrees of freedom available for the August 1 forecast, but only six for the September 1 forecast. The winter wheat analysis had the fewest number of degrees of freedom (three available for the May 1 forecast), since only six years of data were used. (See notes: Tables 6, 7 and 8 for outliers removed from the models.)

Evaluating Differences in Precipitation Data Due to Editing

A station by station data comparison was performed for all States. This portion of the analysis compared data from automated stations that appeared in both the NCDC and CAC networks. Approximately 71% of all stations in all States that appeared in the March through December 1992 Dial-Up Service data sets were successfully matched by NCDC Asheville station reference number with NCDC stations. Differences were calculated as NCDC minus CAC values for matching stations and a paired t-test was performed for each of 47 States in the continental United States. Rhode Island was dropped from the analysis because only one CAC station out of two (Providence) was successfully matched. Four monthly total precipitation terms were examined (March, June, July, and August) which were identified (McCormick, 1993) for use in forecasting models for corn (C), soybeans (S), and winter wheat (W). The mean difference used for the t-test, the maximum absolute difference over all stations in the State and the mean relative difference were calculated.

The final portion of this analysis entailed making 1992 forecasts using different data sets for model building and model input. Proxy CAC data series were constructed from stations that had matching NCDC station reference numbers over the years from 1980 through 1991 from both the NCDC and CAC data sets. Four model input scenarios were examined:

1. All NCDC data (1980-1991), were used to build parameters and all 1992 NCDC data were used to calculate model inputs - denoted by NCDC($NCDC_a$).

2. All NCDC data (1980-1991), were used to build parameters and 1992 NCDC data, matched to CAC stations, were used to calculate model inputs - denoted by NCDC($NCDC_m$).

3. All NCDC data (1980-1991), were used to build parameters, and 1992 CAC station data, matched to NCDC stations, were used to calculate model inputs - denoted by NCDC(CAC_m).

4. "Proxy" CAC data from NCDC stations matched to CAC stations, (1980-1991), were used to build parameters and 1992 CAC matched station data were used to calculate model inputs - denoted by CAC(CAC_m).

In general, the differences between the first (NCDC($NCDC_a$)) and second (NCDC($NCDC_m$)) forecasts represent the effect of using the different station densities for forecasting. The differences between the second and third (NCDC(CAC_m)) forecasts represent the effect of using edited data rather than original NCDC data for forecasting. Then, the differences between the third and fourth (CAC(CAC_m)) forecasts represent the effect of using the "proxy" CAC data for model building.

RESULTS

Differences Due to Density

Mean differences between the NCDC and "proxy" CAC monthly precipitation totals for the regions (over years) were calculated. The difference was calculated as:

$$\text{Difference} = \text{NCDC} - \text{CAC}.$$

Tables 3, 4, and 5 present the results of the paired comparison t-tests between NCDC and CAC-based terms for each crop. This illustrates that corn and winter wheat mean values are significantly different at the α =0.05, 0.01 levels respectively.

Tables 6, 7 and 8 present the prediction intervals and R_a^2 for each crop comparing the official linear or quadratic model using survey data only versus the addition of a NCDC-based or CAC-based optimal monthly precipitation term. In each table, the prediction intervals relate to the years with minimum, median, and maximum regional crop yields.

The prediction intervals and R_a^2 values from the NCDC and CAC models tend to be very similar for corn and soybeans. Surprisingly, the wheat CAC model has smaller prediction intervals than the NCDC model, even though the CAC model is based on less precipitation data. The forecasting models containing a precipitation term provide smaller prediction intervals for corn and winter wheat than the official models. The R_a^2 and PI values for soybeans are similar for precipitation and official models.

Differences Due to Editing

Table 9 presents a list of States that had significant differences in reported total monthly precipitation between 1992 Climatic Analysis Center Dial-up Service station data and National Climatic Data Center station data for matching stations. Eleven out of 47 States showed significant differences for at least one specific term. Most mean differences are smaller than 0.50, but there are cases where high

Table 3: Paired Comparison t-test Corn

YEAR	JUNE NCDC (in).	JUNE CAC (in).	JUNE DIFFERENCE (NCDC-CAC)	JULY NCDC (in).	JULY CAC (in).	JULY DIFFERENCE (NCDC-CAC)
1980	4.27	4.06	0.21	2.81	2.79	0.02
1981	4.90	4.79	0.11	4.93	4.80	0.13
1982	3.43	3.29	0.14	4.60	4.52	0.08
1983	4.47	4.32	0.15	2.57	2.48	0.09
1984	4.75	4.56	0.19	3.12	3.07	0.05
1985	3.30	3.07	0.23	3.06	3.09	-0.03
1986	4.40	4.56	-0.16	4.91	4.68	0.23
1987	3.07	2.92	0.15	4.61	4.36	0.25
1988	1.36	1.40	-0.04	2.72	2.67	0.05
1989	3.36	3.18	0.18	3.54	3.49	0.05
1990	5.99	5.82	0.17	4.76	4.72	0.04
1991	3.14	3.42	-0.28	2.71	2.63	0.08
Mean	3.70	3.61	0.09**	3.87	3.78	0.09*

* Significant at the $\alpha = 0.10$ level of significance.

** Significant at the $\alpha = 0.05$ level of significance.

Table 4: Paired Comparison t-test Soybeans

YEAR	JULY NCDC (in).	JULY CAC (in).	JULY DIFFERENCE (NCDC-CAC)	AUGUST NCDC (in).	AUGUST CAC (in).	AUGUST DIFFERENCE (NCDC-CAC)
1980	2.80	2.61	0.19	†		
1981	5.38	5.28	0.10			
1982	4.47	4.03	0.44			
1983	2.39	2.28	0.11	2.25	2.34	-0.09
1984	3.23	3.05	0.18	2.00	2.26	-0.26
1985	3.09	3.26	-0.17	4.57	4.64	-0.07
1986	4.90	4.39	0.51	3.01	3.12	-0.11
1987	4.74	4.43	0.31	4.85	4.24	0.61
1988	3.09	3.50	-0.41	3.25	3.66	-0.41
1989	3.86	4.37	-0.51	3.62	3.50	0.12
1990	4.81	4.61	0.20	3.81	3.33	0.48
1991	2.69	2.65	0.04	2.95	3.03	-0.08
Mean	3.79	3.71	0.08	3.37	3.35	0.02

Neither are significantly different.

† 1980-1982 data are not used to make the September forecast for soybeans.

Table 5: Paired Comparison t-test Wheat

YEAR	MARCH + AUGUST NCDC (in).	CAC (in).	DIFFERENCE (NCDC-CAC)
1983	4.74	4.62	0.12
1985	4.24	3.73	0.51
1986	3.76	3.32	0.44
1987	5.94	5.29	0.65
1988	5.02	4.77	0.25
1989	4.05	3.67	0.38
1991	4.31	3.90	0.41
Mean	4.58	4.19	0.39*

* Significant at the $\alpha = 0.01$ level of significance.

† 1980-1982; 1984 and 1990; (1987 is the outlier year) data are not used to make the May forecast for winter wheat.

Table 6: Corn Model Forecast Results

Model	R_a^2	Prediction Intervals min	med	max
AUGUST:				
OFFICIAL	.88	6.5	5.3	5.3
JUL NCDC	.92	5.4	4.4	4.6
JUL CAC	.92	5.4	4.2	4.4
SEPTEMBER:				
OFFICIAL	.93	4.8	4.0	4.1
JUN NCDC	.98	2.5	2.1	2.1
JUN CAC	.98	2.7	2.3	2.3

MINIMUM:	1983	82.0 bushels
MEDIAN:	1989	120.4 bushels
MAXIMUM:	1986	125.2 bushels

Note: Every model is reported with observations from outlier year 1988 removed.

Table 7: Soybean Model Forecast Results

Model	R_a^2	Prediction Intervals min	med	max
AUGUST:				
OFFICIAL	.63	2.9	2.4	2.4
JUL NCDC	.64	3.0	2.5	2.4
JUL CAC	.62	3.1	2.6	2.5
SEPTEMBER:				
OFFICIAL	.93	1.4	1.4	1.3
AUG NCDC	.92	1.5	1.5	1.4
AUG CAC	.92	1.6	1.5	1.4

MINIMUM:	1988	27.8 bushels
MEDIAN:	1983	29.0 bushels
MAXIMUM:	1990	37.4 bushels

Note: No outliers were detected.

Table 8: Winter Wheat Model Forecast Results

Model	R_a^2	Prediction Intervals min	med	max
MAY:				
OFFICIAL	.88	1.7	1.5	1.7
AUG+MAR NCDC	.94	1.2	1.1	1.3
AUG+MAR CAC	.97	0.9	0.8	1.0

MINIMUM:	1989	32.9 bushels
MEDIAN:	1985	37.5 bushels
MAXIMUM:	1983	42.4 bushels

Note: Every model is reported with observations from outlier year 1987 removed.

TABLE 9: STATES WITH SIGNIFICANT MEAN DIFFERENCES ($\alpha = 0.10$)
BETWEEN 1992 NCDC AND CAC MONTHLY PRECIPITATION TERMS
OVER MATCHING STATIONS

STATE	# STATIONS matched(actual)	% MATCHED	TERM	CROP C/S/W	MEAN CAC	MEAN NCDC	MEAN DIFFS.[1]	ABS. MAX. DIFFS.	MEAN REL. DIFFS.
AL	7(10)	70	AUG	S	5.38	5.78	0.40	1.25*	0.09
AZ	8(11)	73	MAR	W	1.68	2.04	0.53	1.92	0.26
CA	21(47)	45	MAR	W	2.74	3.13	0.58	3.32*	0.13
KY	7(7)	100	MAR	W	4.86	5.17	0.16	0.38	0.03
MI	15(20)	75	MAR	W	2.11	2.14	0.16	1.00	0.08
			JUN	C	1.43	1.87	0.41	2.32	0.20
MN	8(10)	80	MAR	W	1.33	1.35	0.23	0.06	0.02
MT	13(13)	100	MAR	W	0.55	0.63	0.11	0.62	0.25
			JUL	C	1.57	2.46	1.00	4.29*	0.27
NM	7(12)	58	MAR	W	0.48	0.65	0.06	0.11	0.24
OR	11(12)	92	JUN	C	1.38	1.61	0.08	0.40	0.05
VA	8(13)	62	AUG	S	3.90	6.01	0.32	0.87	0.08
WV	10(10)	100	JUN	C	2.96	3.06	0.09	0.36	0.03

matched: 115/165 stations (70%)

* Individual station data presented in Table 10.

1 All differences are based on matching station positive data. Some stations had missing reports which were not included in the calculations.

TABLE 10: WORST CASES BY CROP AFFECTED

CROP REGION:

CORN: **State: Montana** **Month: July** **mean difference: 1.00** **absolute maximum: 4.29**

STATIONS	STATION#	STATION TYPE	NCDC	CAC	DIFFERENCE
Cut Bank	2173	Secondary	4.67	0.38	4.29*
Havre	3996	Primary	4.30	1.33	2.97
Bozeman	1050	Secondary	2.81	1.23	1.58

SOYBEANS: **State: Alabama** **Month: August** **mean difference: 0.40** **absolute maximum: 1.25**

STATIONS	STATION#	STATION TYPE	NCDC	CAC	DIFFERENCE
Huntsville	4064	Secondary	3.38	2.13	1.25*
Anniston	0272	Secondary	6.47	5.56	0.91
Birmingham	0831	Primary	7.43	6.99	0.44

WHEAT: **State: California** **Month: March** **mean difference: 0.58** **absolute maximum: 3.32**

STATIONS	STATION#	STATION TYPE	NCDC	CAC	DIFFERENCE
Santa Anna	7888	Secondary	5.48	2.16	3.32*
Long Beach	5085	Primary	5.29	2.69	2.60
Riverside	7470	Primary	3.52	2.68	0.84
San Diego	7740	Primary	4.42	1.76	2.66

Note: Of the "worst cases", the ratio of the number of primary stations to the total number of stations is 0.50. CAC noted that the three stations in Montana are "typically suspicious". The rest are usually always "good reporters" with the exception of the station located in Long Beach, California.

TABLE 11: 1992 EARLY SEASON FORECASTS

August 1 Corn: Y_{92} = **134.1**

	$\hat{Y}_{92}$	$\hat{\epsilon}_{92}$	**95 % PI**
1. NCDC($NCDC_a$)*	123.0	11.1	(114.3,131.7)
2. NCDC($NCDC_m$)	120.1	14.0	(111.4,128.8)
3. NCDC(CAC_m)	119.5	14.6	(110.8,128.2)
4. CAC(CAC_m)†	134.8	-0.7	(126.4,143.2)

August 1 Soybeans: Y_{92} = **40.2**

	$\hat{Y}_{92}$	$\hat{\epsilon}_{92}$	**95 % PI**
1. NCDC($NCDC_a$)	43.0	-2.8	(38.3, 47.7)
2. NCDC($NCDC_m$)	41.9	-1.7	(37.2, 46.6)
3. NCDC(CAC_m)	41.8	-1.6	(37.1, 46.5)
4. CAC(CAC_m)	39.4	0.8	(34.6, 44.2)

May 1 Winter Wheat Y_{92} = **36.1**

	$\hat{Y}_{92}$	$\hat{\epsilon}_{92}$	**95% PI**
1. NCDC($NCDC_a$)	36.8	-0.7	(34.9, 38.7)
2. NCDC($NCDC_m$)	37.1	-1.0	(35.2, 39.0)
3. NCDC(CAC_m)	36.9	-0.8	(35.0, 38.8)
4. CAC(CAC_m)	36.9	-0.8	(35.4, 38.4)

XXXX($XXXX_{a/m}$): first term indicates model building source, and second term indicates 1992 model input source.

a = all stations used,

m = matched stations used,

Y_{92} = actual weighted aggregated yield for OY States (listed previously) in 1992,

$\hat{Y}_{92}$ = model predicted yield, and

$\hat{\epsilon}_{92}$ = residual error between Y_{92} and $\hat{Y}_{92}$.

absolute maximum differences have occurred. Relative differences of State monthly terms are compared in Appendix B.

Differences in predictions between NCDC($NCDC_a$) and NCDC($NCDC_m$) show differences due to density; between NCDC($NCDC_m$) and NCDC(CAC_m) show differences due to editing.
Table 10 shows that each major crop region includes at least one State that reported high absolute maximum differences between data sources by stations during 1992. It also shows that primary (Synoptic) and secondary (Airways) stations are both reporting incomplete or no precipitation data. In general, this would not be expected from the primary stations, but occasionally stations may get shut down or undergo staff cuts.

Table 11 concentrates on the first forecast period for each crop, since the effects of precipitation are largest for this period. Comparing forecast results from scenarios 1, 2, and 3 for each crop, the results of NCDC($NCDC_m$) show differences due to density; between NCDC($NCDC_m$) and NCDC(CAC_m) show differences due to editing. Differences due to editing tend to be smaller. The proxy CAC data series (scenario 4) performs well as a model-building source for 1992 forecasts and provides smaller residual errors between the actual Y_{92} and predicted values $\hat{Y}_{92}$. The result corresponds to previous results in Tables 6, 7, and 8 examining NCDC and "proxy" CAC-based models.

CONCLUSIONS

The density of CAC data appears to be adequate for making regional yield forecasts in spite of the fact that the CAC network of stations is based on a reduced number of stations that only provide about 70% coverage. More specifically:

1. Regional precipitation values based on the CAC density are significantly different than NCDC values for corn and winter wheat at the $\alpha = 0.10$ level of significance. Corn regional mean differences for June and July are 0.09 inches for both months. The mean precipitation difference for the May winter wheat forecast term is larger, at 0.39 inches. Soybean regional mean differences are 0.09 and 0.02 inches for July and August, respectively. Neither is significantly different. The corn and soybean differences are relatively small and, if significantly different, have no practical importance. The winter wheat regional mean difference (0.39) represents approximately 8% of the average August and March total, and indicates density may have more of an effect over the winter wheat States.

2. Regional crop yield models developed using "proxy" CAC-based terms were just as accurate as those based on NCDC terms. Prediction intervals and R_a^2 values from the CAC models were comparable to values produced by the NCDC models for corn and soybeans. The winter wheat CAC

prediction intervals were actually about 25% smaller than the NCDC intervals. CAC models capture the impact of precipitation on final yield comparably to NCDC models even though CAC values are based on only about one-tenth the number of stations as NCDC values and only provide coverage on approximately 70% of crop harvested acres.

3. Small mean differences attributed to editing were found between the original CAC and final NCDC values at the State level over matching stations in 1992. However, there were cases where large absolute maximum differences occurred in monthly data reports for individual stations that reported to both the CAC and NCDC. Examples of these discrepancies have been reported to CAC and are being investigated. Differences in the precipitation terms due to editing have minimal effects on regional level forecasts, based on one year of analysis.

RECOMMENDATIONS

Based upon these findings, the following recommendations are made:

(1) CAC data are a relatively reliable current year input to crop yield forecasting models. It can be used for corn and soybean forecasts, but with caution for winter wheat forecasts. In any case, continual data quality evaluation is advised as editing procedures evolve.

(2) While evaluating data editing procedures, NASS should continue to update the NCDC historic precipitation data base annually until a sizeable CAC data series can be fully evaluated.

(3) A full evaluation of CAC data and editing procedures requires that NASS continue to maintain the CAC monthly data series.

(4) Preserving existing inter-Agency cooperative agreements, and efforts (JAWF) is desired. NASS should continue to work with the CAC to improve the quality of near real-time data to minimize changes due to editing through further evaluation and monitoring of data editing procedures each year. Continued analysis is emphasized.

BIBLIOGRAPHY

Arends, W., Barrowman, D.W., Bridge, R., Buche, J., Frank, D., Klugh, B., Matthews, W., Witzig, J., (1983) "Scope and Methods," Miscellaneous Publication Number 1308, SRS, U.S. Department of Agriculture.

Belsley, D. A., Kuh, E., Welsh, R.E., (1980), Regression Diagnostics, John Wiley & Sons.

Birkett, T. R., (1990) "The New Objective Yield Models for Corn and Soybeans," SMB Staff Report Number SMB 90-02, U.S. Department of Agriculture.

Birkett, T. R., (1993) "Yield Models for Corn and Soybeans Based on Survey Data," U.S. Department of Agriculture, Proceedings, International Conference on Establishment Surveys, Buffalo, New York.

Draper, N.R., Smith, H., (1981), Applied Regression Analysis, John Wiley & Sons Second Edition.

Kaiser, M. and Sebaugh, J. L. (1984), "Methods for the Evaluation of Real-Time Weather Data for Use in Crop Yield Models: An Application to North Dakota," AGES840424, U.S. Department of Agriculture in cooperation with the Statistics Department of the University of Missouri.

McCormick, M. D., Birkett, T. R. (1992), "Evaluating the Addition of Weather Data to Survey Data to Forecast Soybean Yields," Research Report No. SRB 92-11, U.S. Department of Agriculture.

McCormick, M. D. (1993), "Using Different Precipitation Terms to Forecast Corn and Soybean Yields," U.S. Department of Agriculture, Proceedings, International Conference on Establishment Surveys, Buffalo, New York.

McCormick, M. D. (1993), "The Effect of Weather Station Density on Crop Yield Forecasts," Conference on Applied Statistics in Agriculture proceedings, Kansas State University, Manhattan, Kansas.

Motha, R.P., Heddinghaus, T.R. (1986), "The Joint Agricultural Weather Facility's Operational Assessment Program," American Meteorological Society.

APPENDIX A

The following procedures are used to aggregate the precipitation and survey data to regional totals:

Precipitation Data

Precipitation variables used in the models represent total precipitation for a particular month at the regional level. The variable is constructed as follows:

$$P_t = \frac{\sum_{s=1}^{S} A_{ts} R_{ts}}{\sum_{s=1}^{S} A_{ts}}, \qquad (1)$$

where

P_t	=	estimated monthly accumulated precipitation for the region for year t,
S	=	the number of States covered,
A_{ts}	=	the acres for harvest for year t, State s, and
R_{ts}	=	the estimated monthly accumulated precipitation for year t, State s, where
A_{tsd}	=	the acres for harvest for year t, State s, district d, and

$$R_{ts} = \frac{\sum_{d=1}^{D_s} A_{tsd} E_{tsd}}{\sum_{d=1}^{D_s} A_{tsd}} ,$$

D_s = the number of districts per State s.

E_{tsd} = the average monthly station accumulated precipitation for year t, State s, district d,

$$E_{tsd} = \frac{1}{W_{tsd}} \sum_{w=1}^{W_{tsd}} U_{tsdw} ,$$

where

W_{tsd} = number of weather stations for year t, State s, district d, and

U_{tsdw} = accumulated monthly precipitation for year t, State s, district d, weather station w.

Survey Data

The construction of the independent variables for the regional regression models for both soybeans and corn is discussed by Birkett (1990, 1993). For soybeans for the month of August, the independent variable (Z_t) is the estimated number of lateral branches per eighteen square feet. For September, the independent variable is the estimated number pods with beans per eighteen square feet. These regional level estimates for soybeans are constructed as follows:

$$Z_t = \frac{\sum_{s=1}^{S} A_{ts} F_{ts}}{\sum_{s=1}^{S} A_{ts}} \qquad (2)$$

where

A_{ts} = the acres for harvest for year t, State s, and

F_{ts} = number of lateral branches per 18 sq. feet year t, State s,

$$F_{ts} = \frac{1}{m_{ts}} \sum_{j=1}^{m_{ts}} B_{tsj} L_{tsj} ,$$

where

m_{ts} = the number of samples in J_{ts} year t, State s,

J_{ts} = the subset of samples classified in maturity categories 2-6 (or 1-6 in southern States), year t, State s,

B_{tsj} = plants per 18 square feet for year t, State s, sample j,

L_{tsj} = lateral branches per plant year t, State s, sample j, (for August), or

= estimated pods with beans per plant per 18 sq. feet, year t, State s, sample j, (for September).

Corn independent variables (Z_t) are more complex and as they are a function of both plant counts and average kernel row length per square foot. C_{ts} is substituted for F_{ts} in equation (2). In August, it is calculated as:

$$C_{ts} = \frac{1}{m_{ts}} \sum_{j=1}^{m_{ts}} (U_{tsj} + V_{tsj})\bar{K}_{tsj} \ ,$$

where

C_{ts} = a function of the number of stalks with ears, the number of ears with kernels, and the average kernel row length per square foot,

U_{tsj} = number of stalks with ears per sq. ft., year t, State s, sample j,

V_{tsj} = number of ears with kernels per sq. ft., year t, State s, sample j, and

$\bar{K}_{tsj}$ = the average kernel row length per ear, year t, State s, sample j.

In September, C_{ts} is calculated as:

$$C_{ts} = \frac{1}{m_{ts}} \sum_{j=1}^{m_{ts}} V_{tsj}\bar{K}_{tsj} \ ,$$

For both forecasts, data are used from the subset of samples in maturity categories 3-6 for year t, State s.

The winter wheat independent variable (Z_t) for may is a function of number of stalks with emerged or late boot heads and weight per head per square foot. Here, H_{ts} is substituted for F_{ts} in equation (2) it is calculated as:

$$H_{ts} = \frac{1}{m_{ts}} \sum_{j=1}^{m_{ts}} Q_{tsj}G_{tsj} \ ,$$

where

H_{ts} = a function of number of stalks with emerged or late boot heads times weight per head per sq. ft. year t, State s,

Q_{tsj} = number of stalks with emerged or late boot heads per sq. ft., year t, State s, sample j, and

G_{tsj} = weight per head year t, State s, sample j.

Yield Data

The regional yield values included in this study were calculated as follows:

$$Y_t = \frac{\sum_{s=1}^{S} A_{ts} Y_{ts}}{\sum_{s=1}^{S} A_{ts}} \ , \qquad (3)$$

where

Y_t = final regional yield for year t, and

Y_{ts} = NASS State yield year t, State s.

APPENDIX B
1992 STATE MONTHLY PRECIPITATION TERM COMPARISONS

STATE	MARCH NCDC	MARCH CAC	MEAN REL. DIFFERENCE	JUNE NCDC	JUNE CAC	MEAN REL. DIFFERENCE
AL	4.2	4.2	0.00	7.1	7.0	0.01
AZ	2.0	1.7	0.26	0.1	0.1	0.00
AR	4.7	3.5	0.06	6.7	6.1	0.14
CA	3.1	2.7	0.13	0.2	0.2	0.11
CO	2.0	2.0	0.03	2.9	2.6	-0.11
CT	3.7	3.7	0.00	5.5	5.5	0.00
DE	4.3	4.2	0.04	2.9	2.9	0.00
FL	3.4	3.0	0.04	12.0	13.2	-0.04
GA	4.3	4.3	0.00	5.8	4.9	0.00
ID	0.4	4.5	-0.10	1.5	1.3	0.18
IL	2.1	2.1	-0.06	1.5	1.5	-0.07
IN	3.5	3.3	0.01	2.4	2.0	0.21
IA	2.2	2.2	-0.02	1.5	1.7	0.16
KS	2.7	2.7	0.00	6.2	6.6	0.06
KY	5.2	4.9	0.03	4.7	4.7	0.01
LA	4.8	4.6	-0.04	7.9	8.5	0.07
ME	3.7	3.3	0.06	4.0	3.4	0.14
MD	4.0	3.9	0.01	2.4	2.4	0.02
MA	3.7	3.2	0.13	4.5	4.4	0.02
MI	2.1	2.1	0.08	1.9	1.4	0.20
MN	1.3	1.4	0.02	3.3	3.3	0.01
MS	4.4	4.6	-0.01	7.0	7.2	-0.02
MO	2.9	2.6	0.17	3.7	3.3	0.20
MT	0.6	0.5	0.25	3.1	3.6	0.13
NE	2.5	2.3	0.06	3.2	3.0	0.04
NV	1.6	1.6	0.02	0.5	0.5	0.12
NH	8.1	5.8	0.17	4.0	4.0	0.00
NJ	3.2	3.1	0.03	3.6	4.1	-0.21
NM	0.7	0.5	0.24	2.0	2.1	-0.21
NY	3.2	2.9	0.07	2.4	2.2	0.05
NC	3.8	3.8	0.00	5.1	5.1	-0.07
ND	0.8	1.0	-0.22	2.9	2.8	0.08
OH	3.4	3.1	0.09	2.4	2.0	-0.01
OK	1.7	1.7	-0.05	6.4	7.0	0.00
OR	1.2	1.0	0.05	1.6	1.4	0.05
PA	3.7	3.6	0.02	1.8	2.1	0.03
RI*	4.0	4.0	0.00	4.6	4.3	0.06
SC	3.8	3.4	0.10	5.5	5.5	0.01
SD	1.2	1.1	0.10	4.6	4.3	0.05
TN	5.4	5.3	0.04	6.2	6.0	0.03
TX	2.7	2.6	-0.01	3.9	3.9	0.02
UT	1.6	1.6	0.04	0.4	0.4	0.02
VT	2.5	2.2	0.11	2.0	2.0	0.01
VA	3.2	3.1	0.04	3.1	2.8	0.11
WA	0.9	1.0	0.15	1.6	1.5	0.11
WV	4.3	4.2	0.02	3.1	3.0	0.03
WI	2.4	2.3	0.04	1.9	2.0	-0.14
WY	1.1	1.1	0.17	1.5	1.3	0.09
U.S.	**3.0**	**2.9**	**0.05**	**3.6**	**3.6**	**0.04**

*Rhode Island has only one matching station located in Providence.

1992 STATE MONTHLY PRECIPITATION TERM COMPARISONS

STATE	JULY NCDC	JULY CAC	MEAN REL. DIFFERENCE	AUGUST NCDC	AUGUST CAC	MEAN REL. DIFFERENCE
AL	6.3	5.9	0.05	5.8	5.4	0.09
AZ	1.9	1.6	-0.30	3.2	3.3	0.11
AR	5.9	5.1	0.16	3.1	2.5	0.08
CA	0.3	0.8	0.20	0.2	0.1	-0.31
CO	1.9	1.7	0.11	2.7	2.7	-0.01
CT	4.2	4.2	-0.01	6.0	6.0	0.01
DE	3.3	3.7	-0.16	4.6	3.3	0.19
FL	3.9	3.1	0.10	8.3	7.1	0.07
GA	6.8	6.3	0.08	6.3	6.3	0.00
ID	0.6	0.5	0.09	0.3	0.4	0.02
IL	7.1	6.8	0.03	1.4	1.4	0.03
IN	8.2	8.0	0.00	2.5	2.3	-0.15
IA	8.4	6.5	0.13	2.3	2.1	0.07
KS	7.5	7.4	0.02	3.3	3.2	0.03
KY	5.8	6.0	0.05	3.9	4.0	-0.01
LA	5.3	4.6	0.14	5.5	5.5	0.02
ME	3.8	3.7	0.02	3.5	3.1	0.13
MD	5.3	4.9	0.08	4.9	4.6	0.04
MA	3.6	3.6	-0.01	4.5	6.2	0.01
MI	4.8	4.2	0.22	2.8	2.5	0.14
MN	3.7	3.7	0.01	3.4	3.4	0.01
MS	4.6	3.9	0.17	7.1	6.0	0.08
MO	7.7	8.0	0.20	2.2	1.7	0.00
MT	2.5	1.6	0.27	1.0	1.1	0.00
NE	5.2	4.9	0.06	3.5	3.5	0.00
NV	0.3	0.4	0.03	0.4	0.6	-0.83
NH	6.6	6.4	0.05	5.3	4.3	0.12
NJ	5.5	5.7	-0.03	5.0	5.0	-0.01
NM	1.7	1.2	0.17	2.5	2.3	0.00
NY	6.3	6.2	0.02	3.4	3.2	0.02
NC	2.9	3.1	-0.03	8.5	8.6	-0.01
ND	2.0	1.8	0.02	1.8	1.6	0.10
OH	9.3	9.6	0.03	3.8	3.6	0.03
OK	3.1	3.4	-0.04	5.3	4.9	0.08
OR	0.7	0.8	-0.20	0.3	0.3	0.13
PA	6.7	6.5	-0.02	3.3	3.1	0.02
RI	3.6	3.4	0.03	6.1	5.6	0.07
SC	3.0	3.1	-0.07	9.1	8.7	0.03
SD	4.7	4.4	0.07	2.3	2.3	0.00
TN	5.6	5.7	0.01	3.2	3.3	0.06
TX	2.5	2.2	0.07	2.1	2.0	-0.01
UT	0.4	0.4	0.00	1.6	1.6	0.00
VT	4.6	3.9	0.15	2.4	2.4	-0.08
VA	4.6	4.1	0.06	6.0	3.9	0.08
WA	1.5	1.1	0.20	1.0	1.2	0.17
WV	6.2	5.9	0.03	3.6	3.3	0.07
WI	4.4	3.8	0.03	2.7	2.7	-0.02
WY	1.8	1.6	0.14	0.8	0.7	0.15
U.S.	**4.3**	**4.1**	**0.05**	**3.6**	**3.4**	**0.02**