DATA QUALITY CONSIDERATION IN USING THE PRISM PROGRAM (Graphics Packet)

A Joint Research Project of Minnegasco and the Minneapolis Energy Office

Timothy S. Dunsworth
Martha J. Hewett
Minneapolis Energy Office
August, 1985

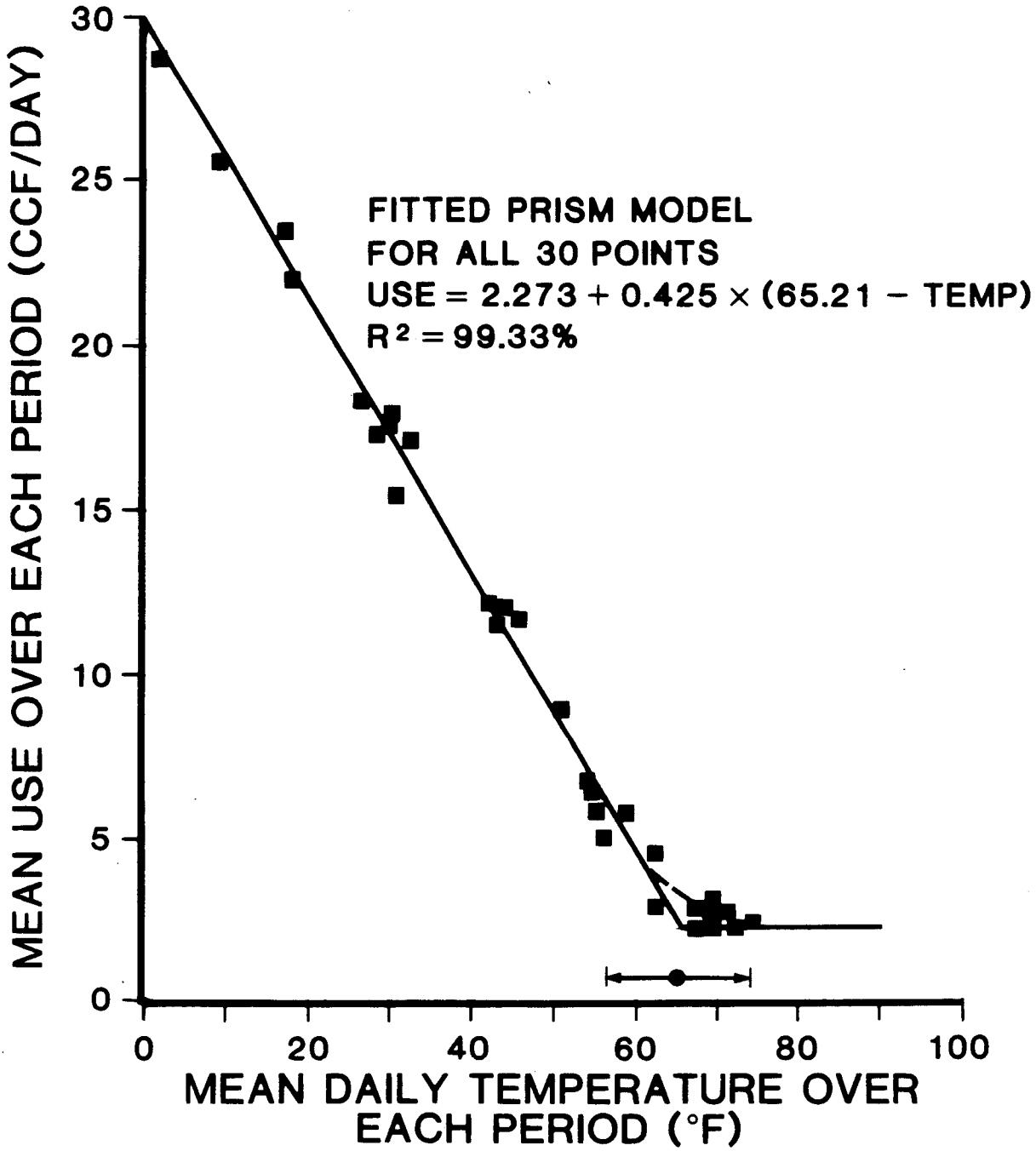
MEO/TR85-5-AM

SOURCES OF GAS CONSUMPTION DATA

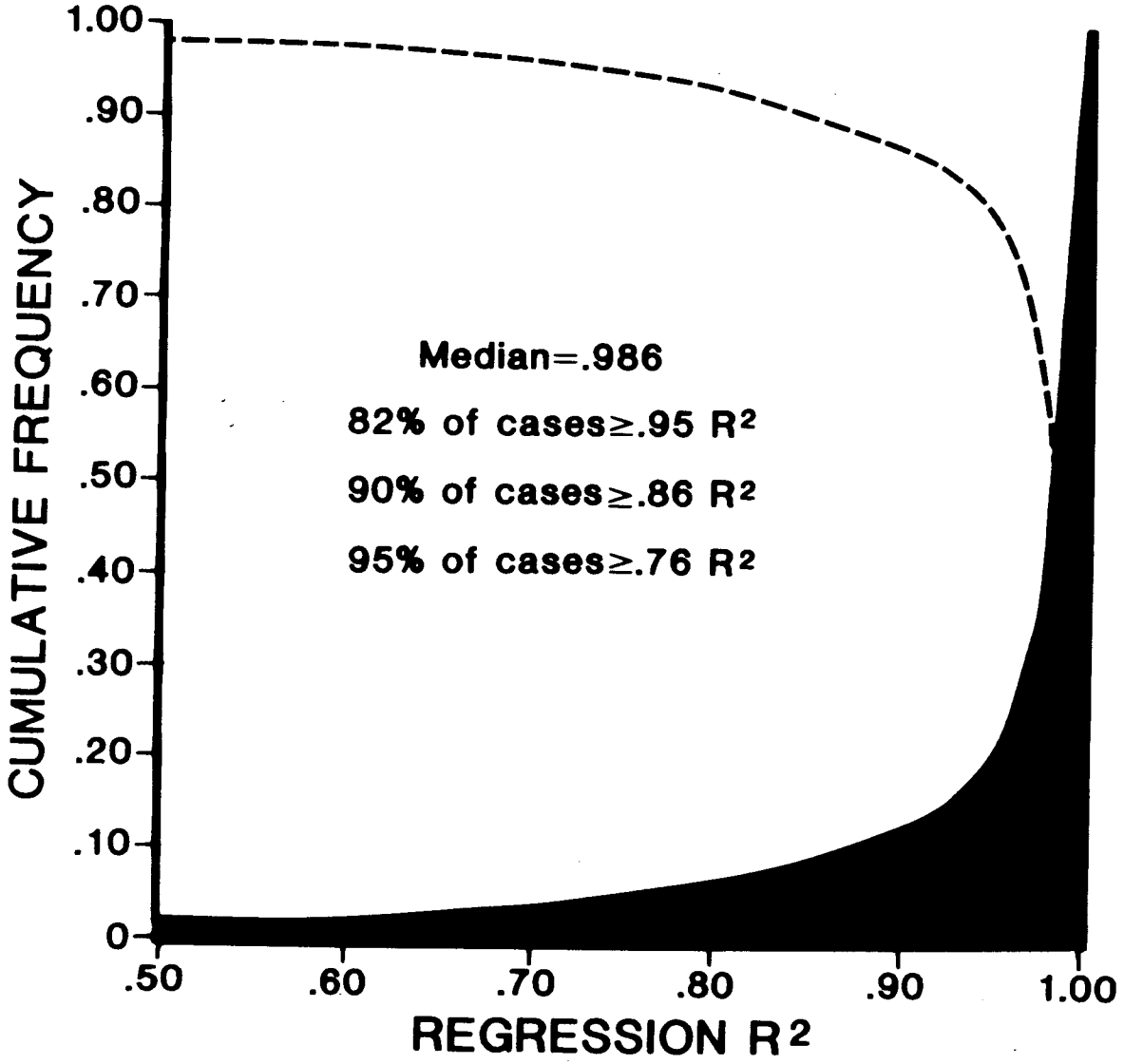
Number of cases	Source Program	Period	Readings/ Months	Number of Datasets
230	Neighborhood Energy Workshops	Pre	Various	197
		Post	Various	211
		(Domestic)	Various	3
1	Energy Bank-Attic	Pre	16/21	32
1	Energy Bank-Furnace	Pre	14/20	36
5	Outdoor Reset Tests	Post	8/11	3
		Post	11/12	3
		Post	9/12	3
		Post	8/12	3
		Post	8/12	3
1	Multifamily Audits	Pre	30/36	279
1	(Artificially Generated)	Pre & Post	27/27	69
239				842

DATA FOR MAIN SIMULATION CASE

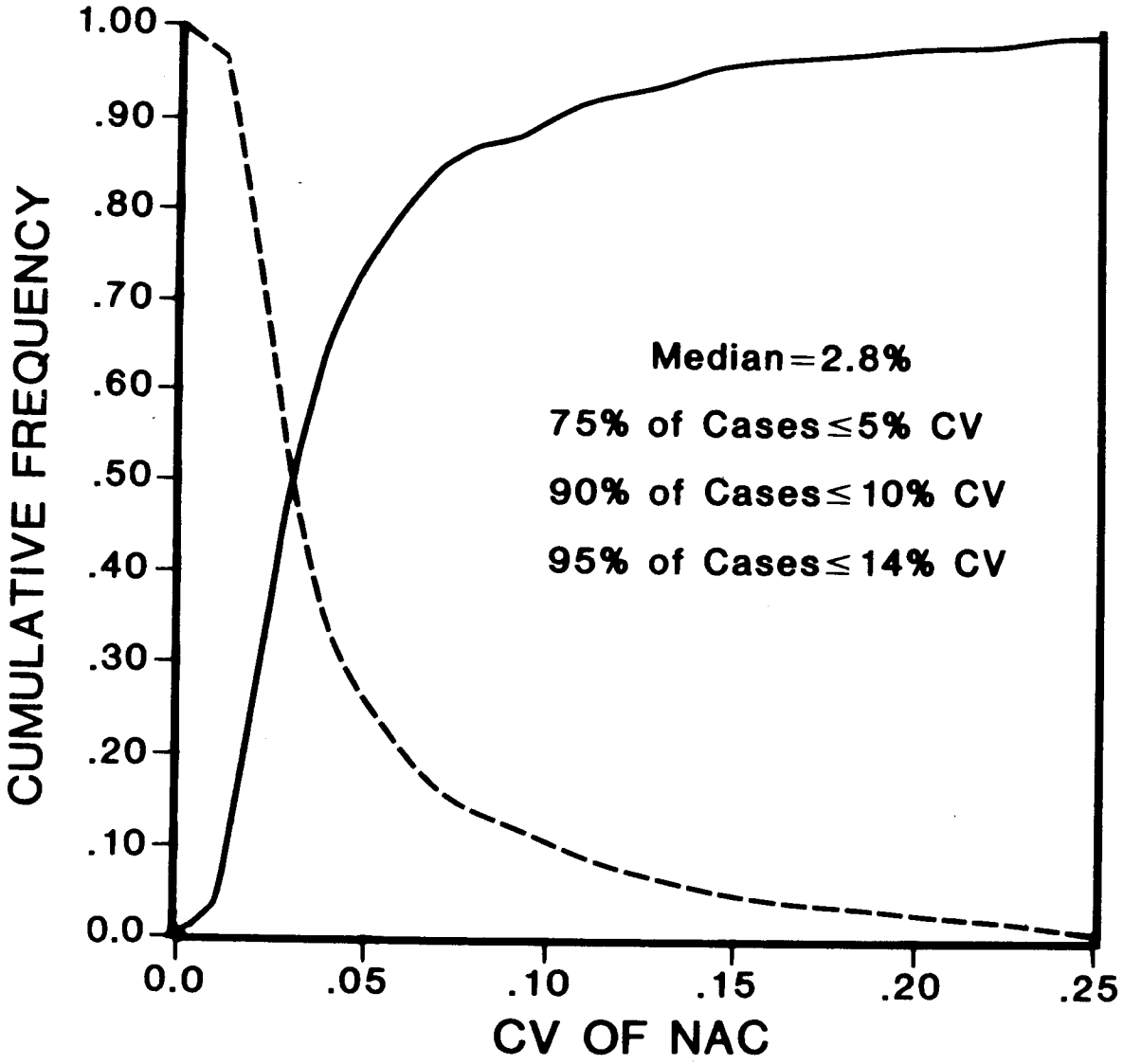
(04-366320, 6/79 THROUGH 6/82)



REGRESSION R – SQUARED DISTRIBUTION FOR 410 N.E.W. CASES



COEFFICIENT OF VARIATION OF NAC FOR 411 N.E.W. CASES



NEIGHBORHOOD ENERGY WORKSHOP RESULTS BY THE 95/5 RULE OF ACCEPTABILITY

	CV of NAC		
	Under 5%	Over 5%	Total
R^2 over 95%	300 f 89.3 r 97.4 c (-- p)	36 f 10.7 r 35.0 c 32.4 p	336 f 81.8 c
R^2 under 95%	8 f 10.7 r 2.6 c 7.2 p	67 f 89.3 r 65.0 c 60.4 p	75 f 18.2 c
Total	308 f 74.9 r	103 f 25.1 r	411 f

f = absolute frequency
r = percent of row total
c = percent of column total
p = percent of all failures

$\chi^2 = 201.79$ $p < .005$

N.E.W. DATA QUALITY VERSUS OUTCOMES AGAINST A 95/5 TEST CRITERION

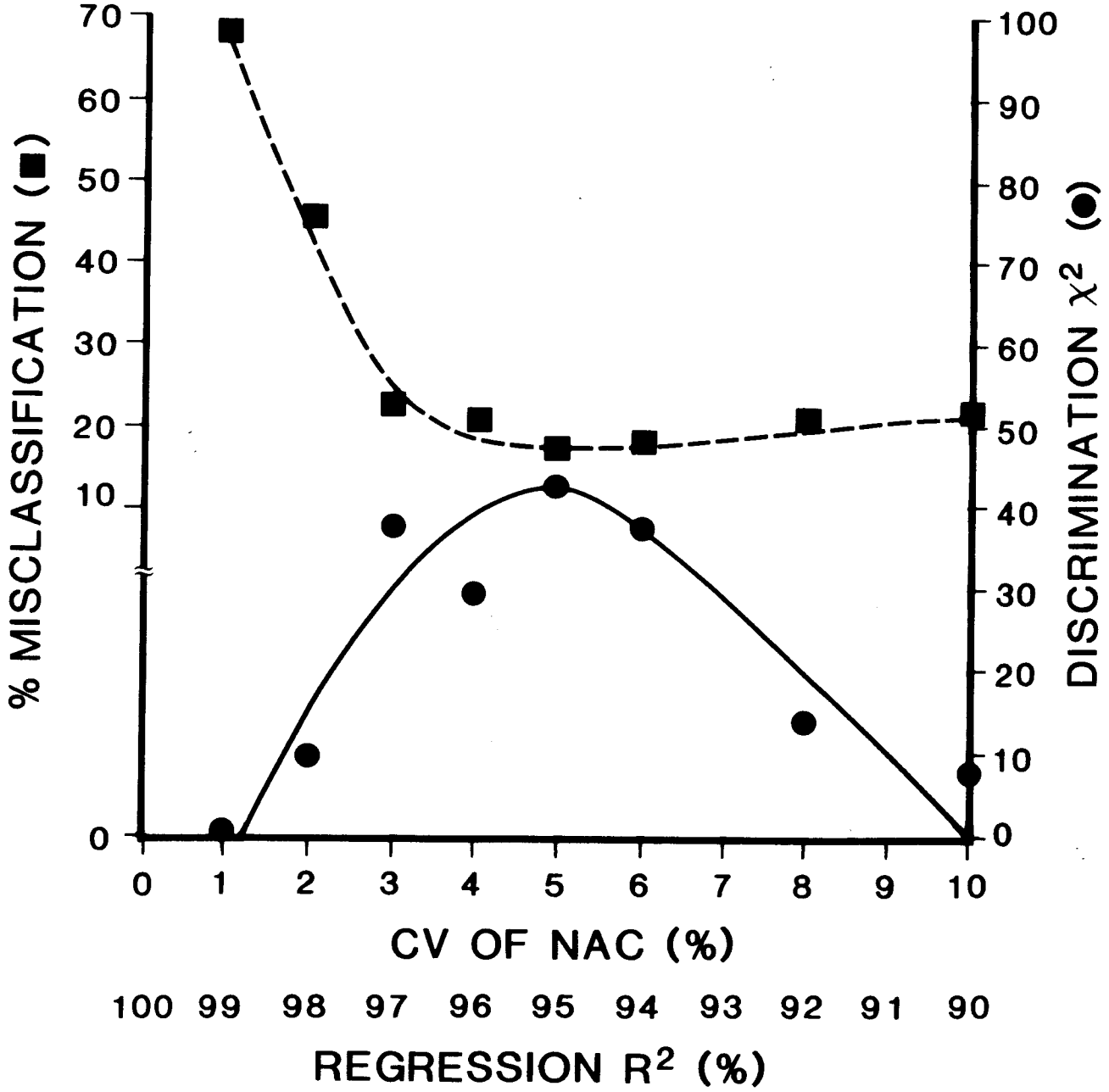
Data quality	Pass R^2 and CV	Fail R^2 and/or CV	Total
<4 readings or grossly inconsistent	(0 by definition)	46 f 100.0 r 28.6 c	46 f 10.0 c
4-5 readings and/or no summer data	2 f 18.2 r 0.7 c	9 f 81.8 r 5.6 c	11 f 2.4 c
marginal	27 f 51.9 r 9.1 c	25 f 48.1 r 15.5 c	52 f 11.3 c
apparently satisfactory	269 f 76.9 r 90.7 c	81 f 23.1 r 50.3 c	350 f 76.3 c
Total	298 f 64.9 r	161 f 35.1 r	459 f

f = absolute frequency

r = percent of row total

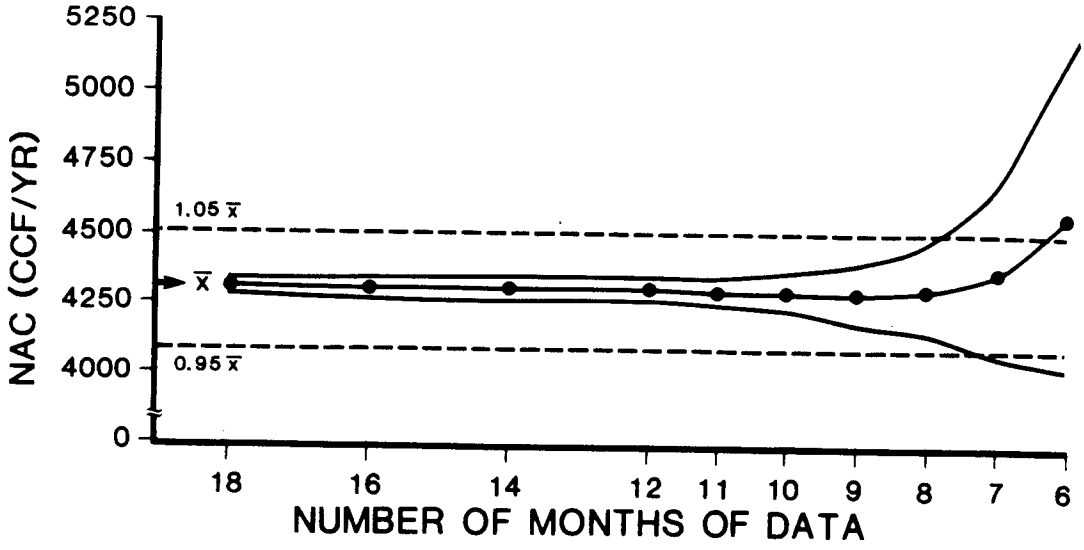
c = percent of column total

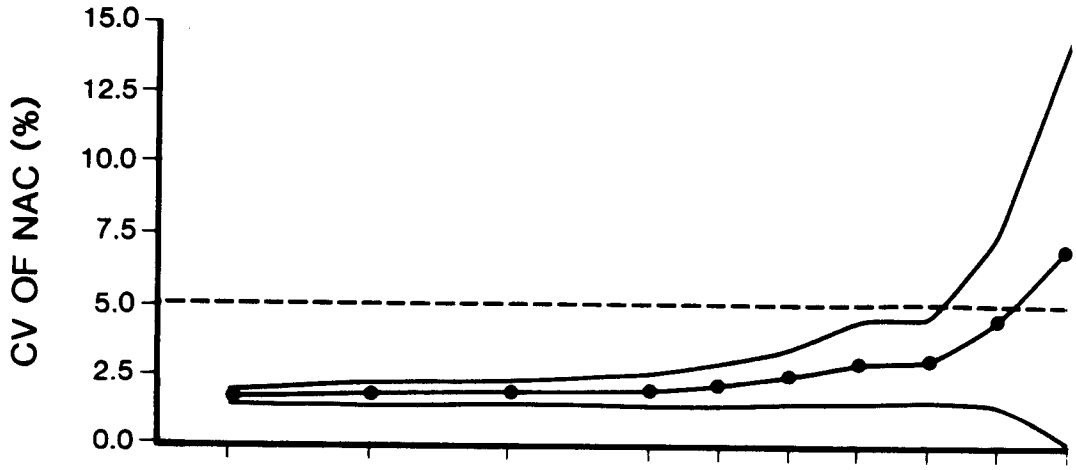
COMPARISON OF VARIOUS SCREENING CRITERIA

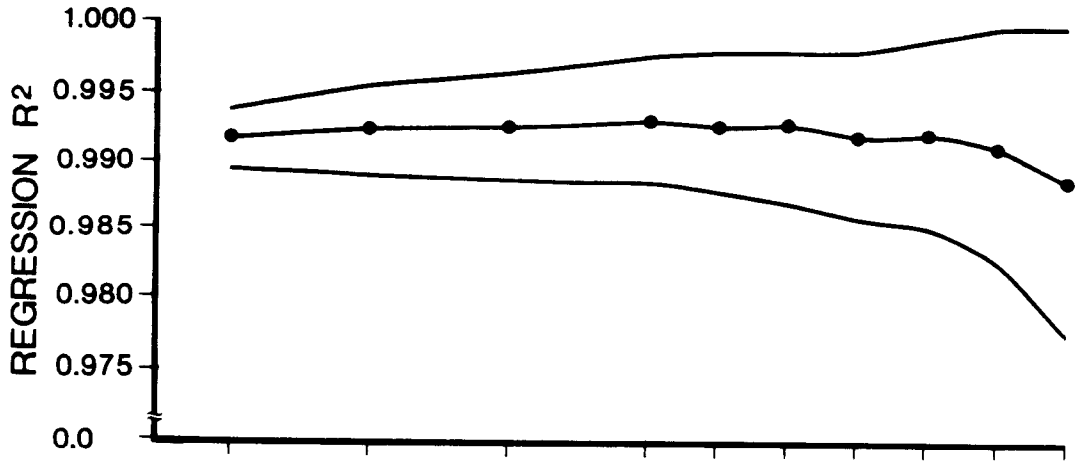


PRISM RESULTS VERSUS NUMBER OF MONTHS OF DATA

(PLOTTED AS MEAN ± STD. DEV. OF OBSERVED VALUES)

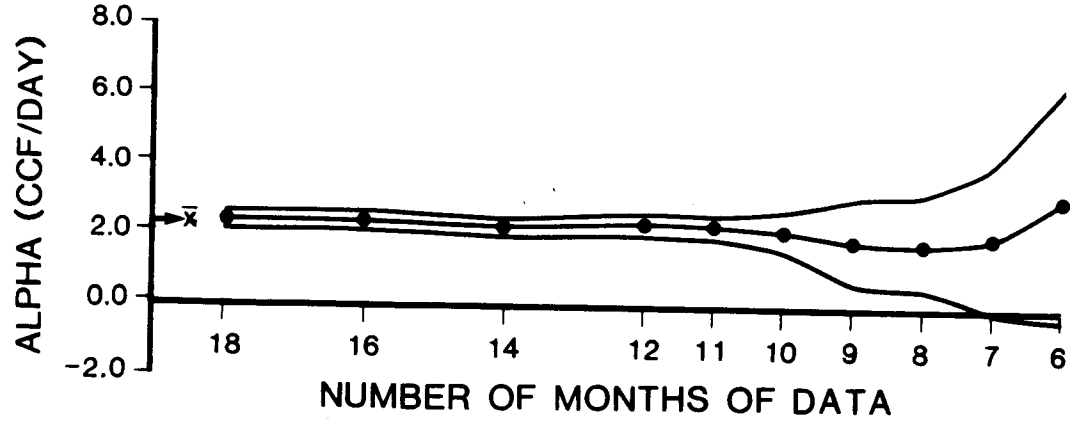


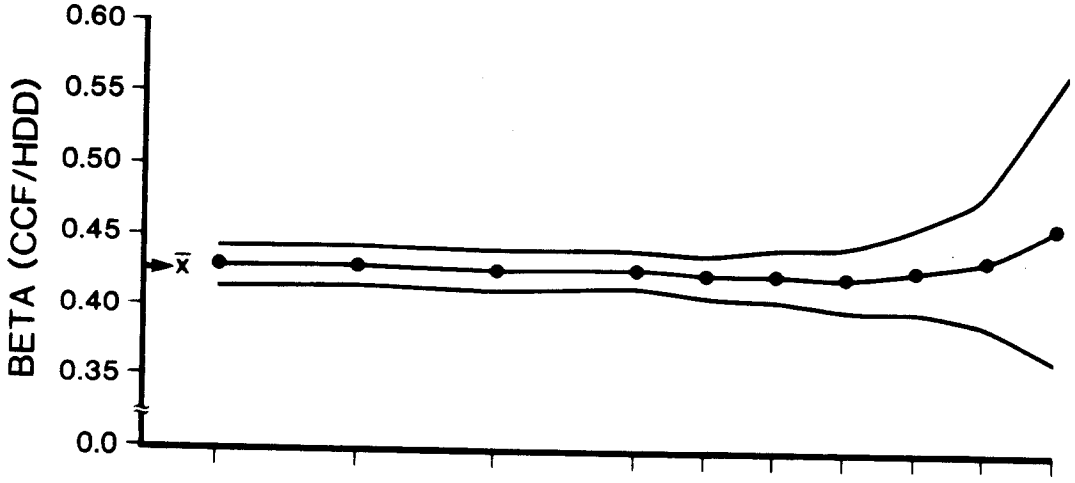


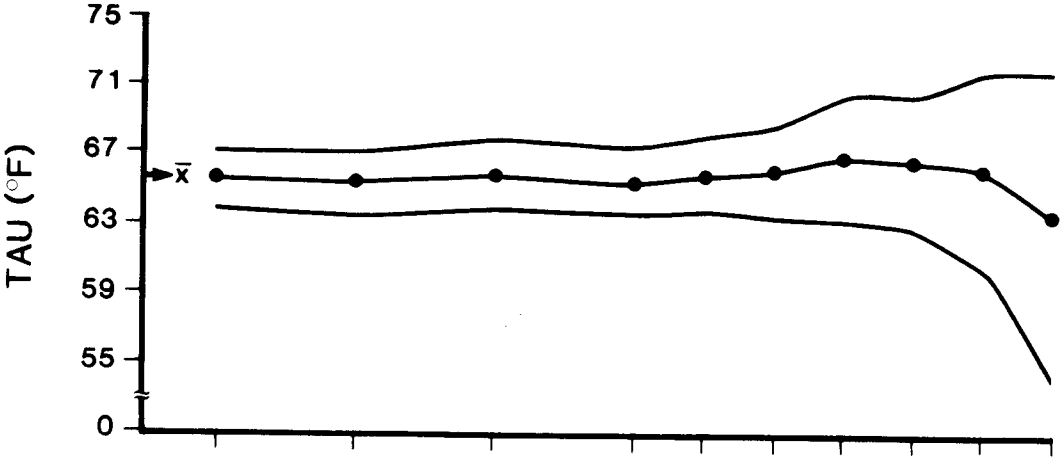


PRISM RESULTS VERSUS NUMBER OF MONTHS OF DATA

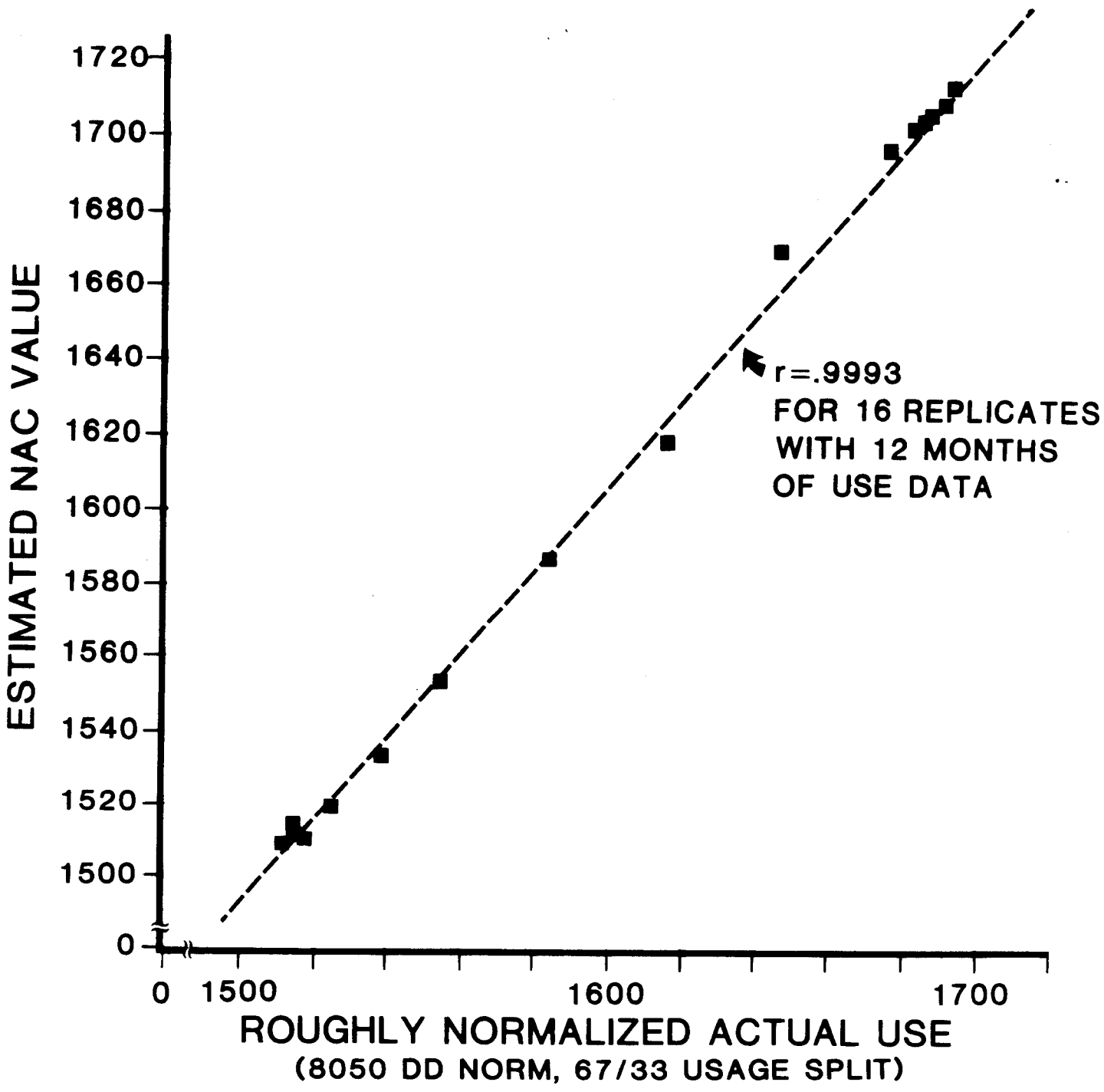
(PLOTTED AS MEAN ± STD. DEV. OF OBSERVED VALUES)





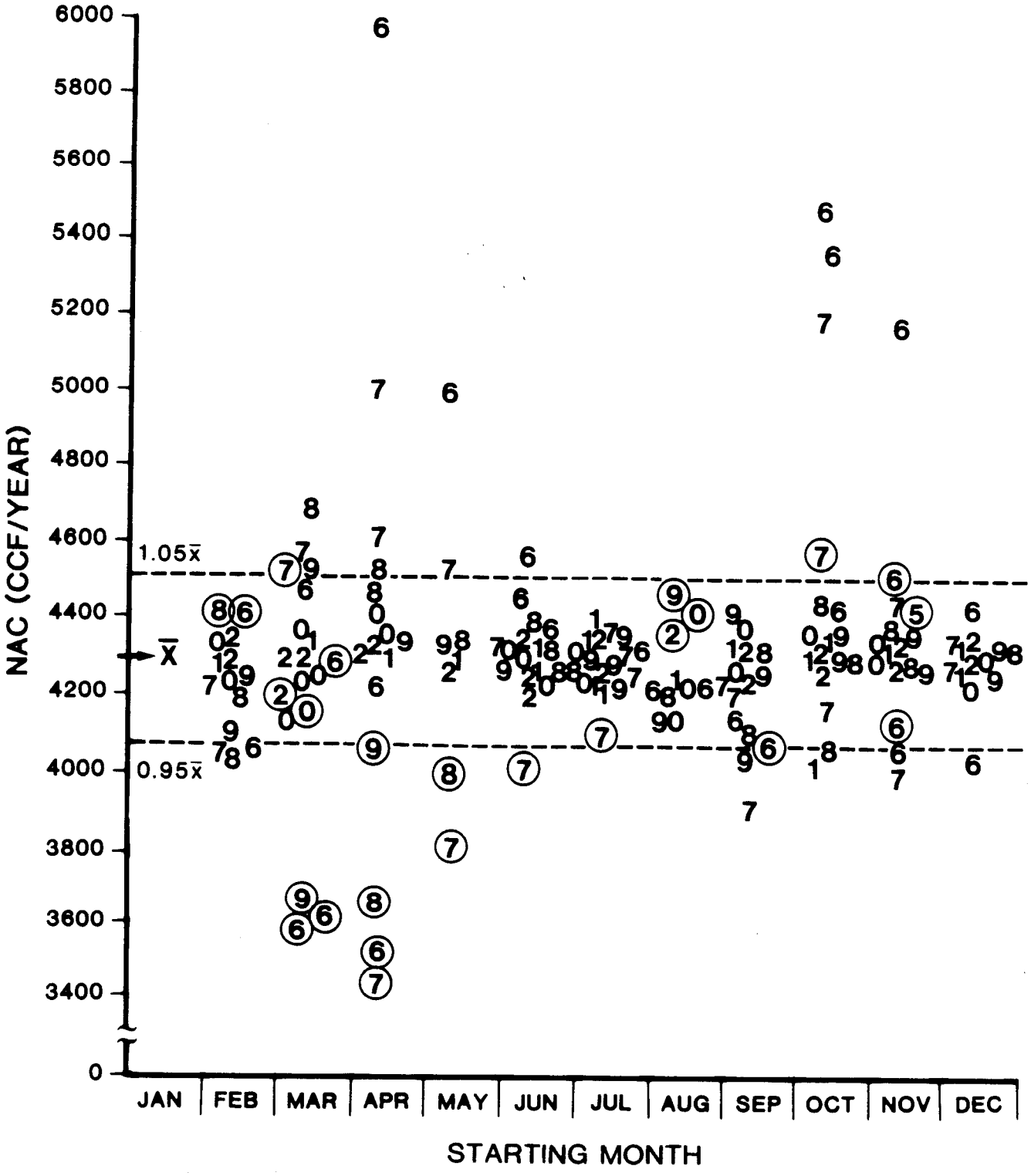


PRISM PERFORMANCE IN TRACKING ACTUAL USE IN A MIXED-MODEL CASE



VARIATION OF NAC OVER TIME IN REPLICATES WITH DIFFERENT NUMBERS OF MONTHS

(EACH POINT MARKED BY THE ONES DIGIT OF THAT NUMBER)

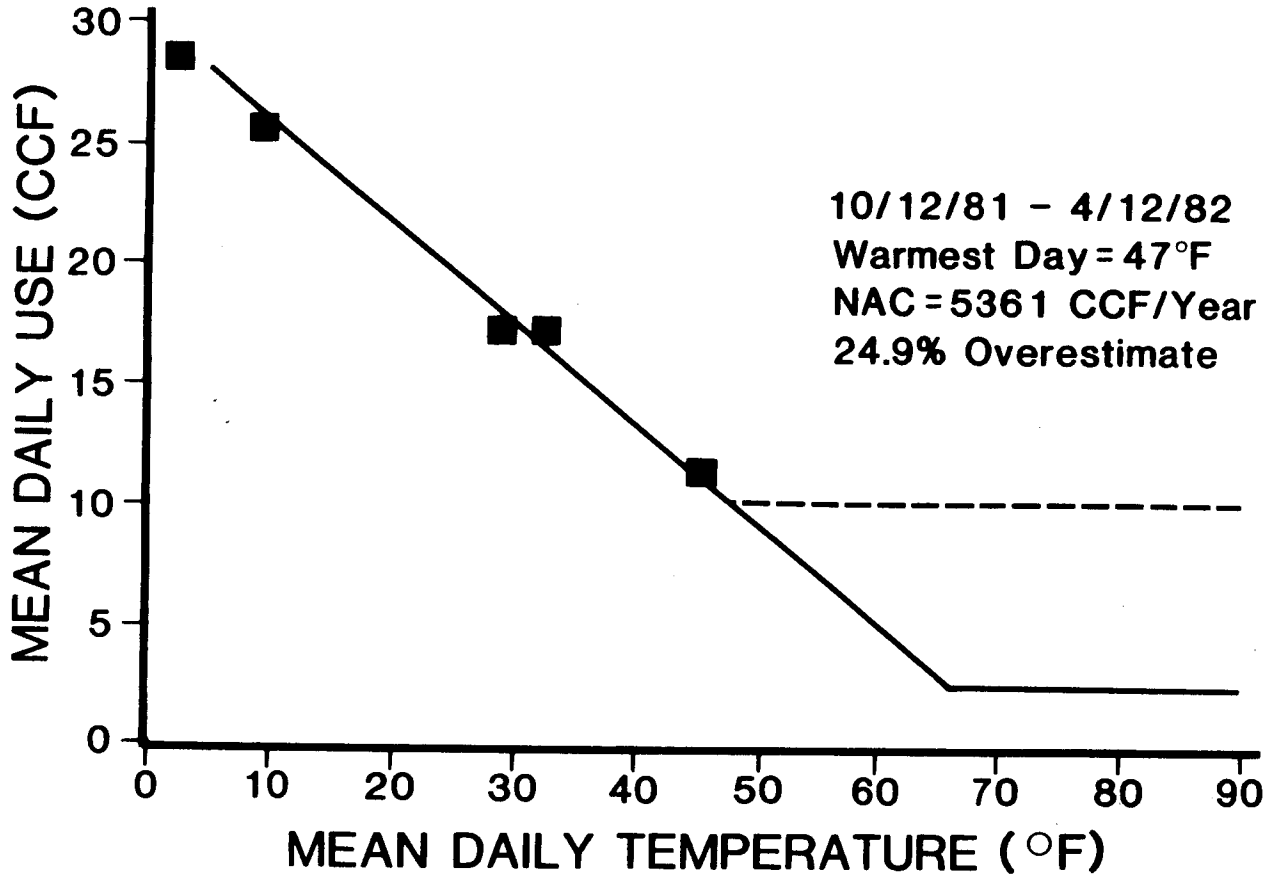


EFFECT OF SUBTRACTING ONE OR TWO MONTHS FROM 12 MONTH REFERENCE PERIODS

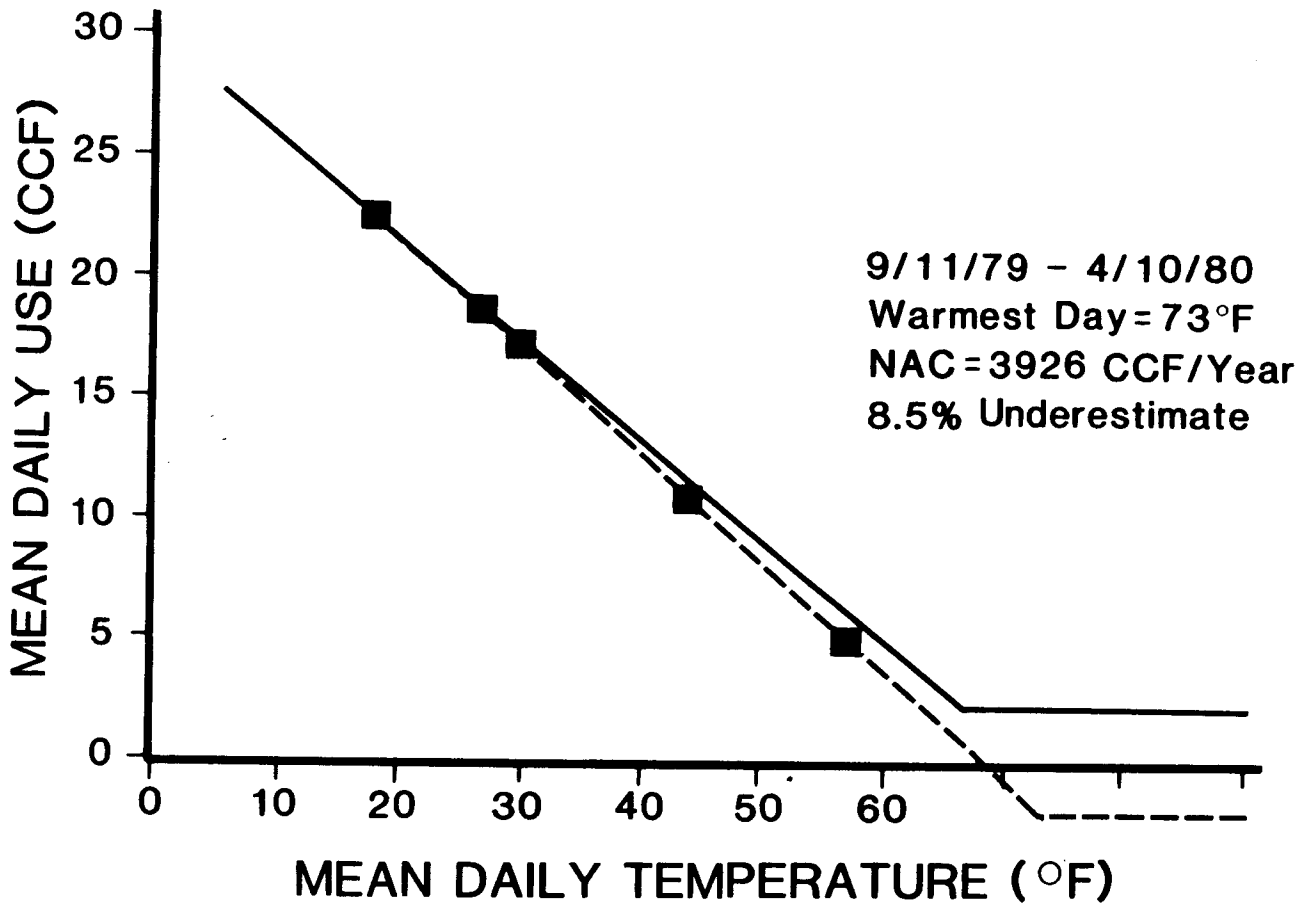
	subtract summer data	subtract spring/fall	subtract winter
n of pairs	15	14	11
mean change	-10.8 ccf	-2.3 ccf	+9.8 ccf
std. dev.	33.4	34.5	53.1
t statistic	-1.253	-0.248	+0.613
p value	.2< p<.4	.5< p<.9	.5< p<.9

Tests are paired sample t tests measuring changes between overlapping periods, and these changes should also be considered against the overall NAC results for 12 month periods of about 4298 ± 43 ccf/year.

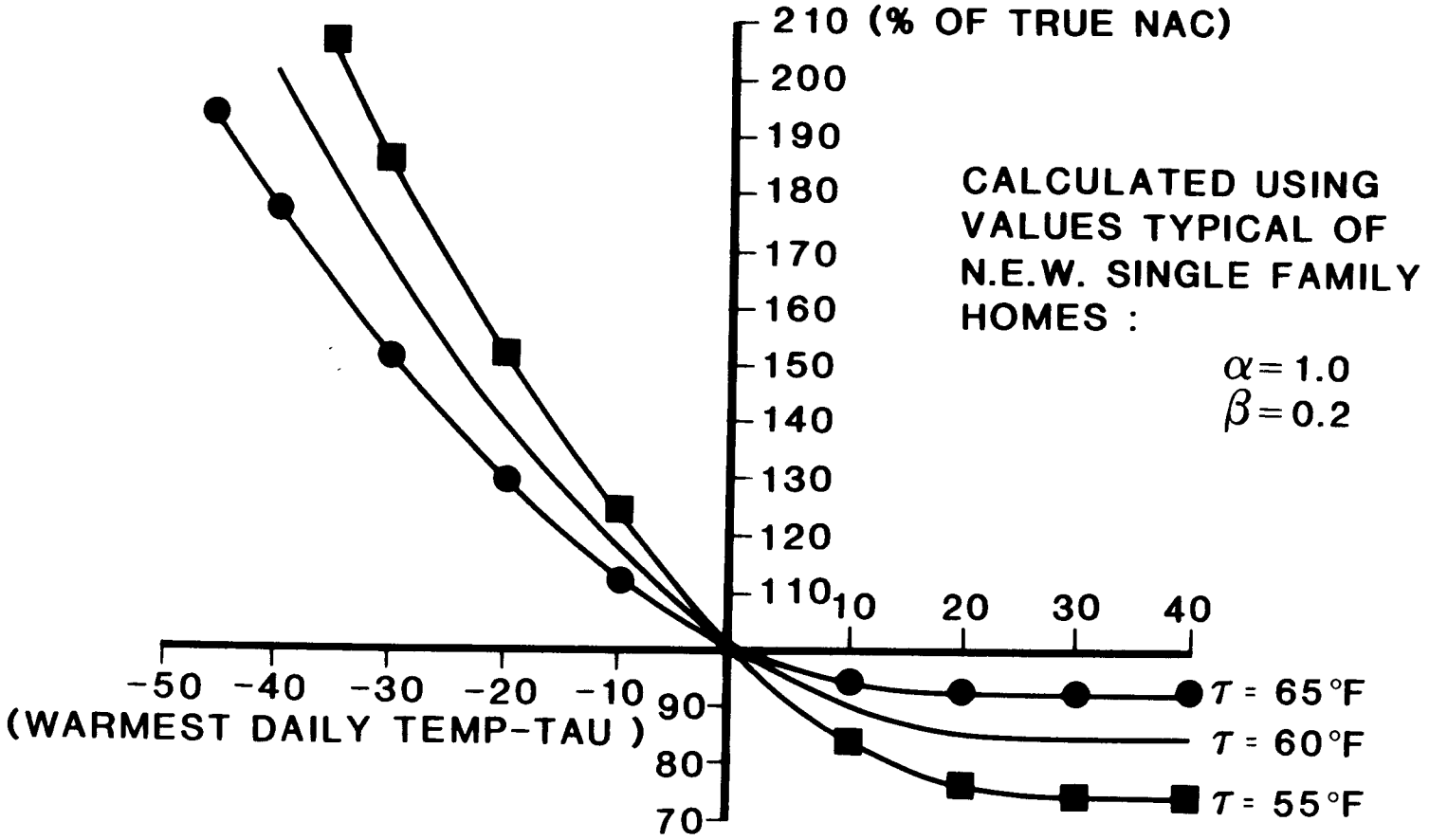
USE VERSUS TEMPERATURE PLOT FOR REPLICATE MISSING SUMMER DATA



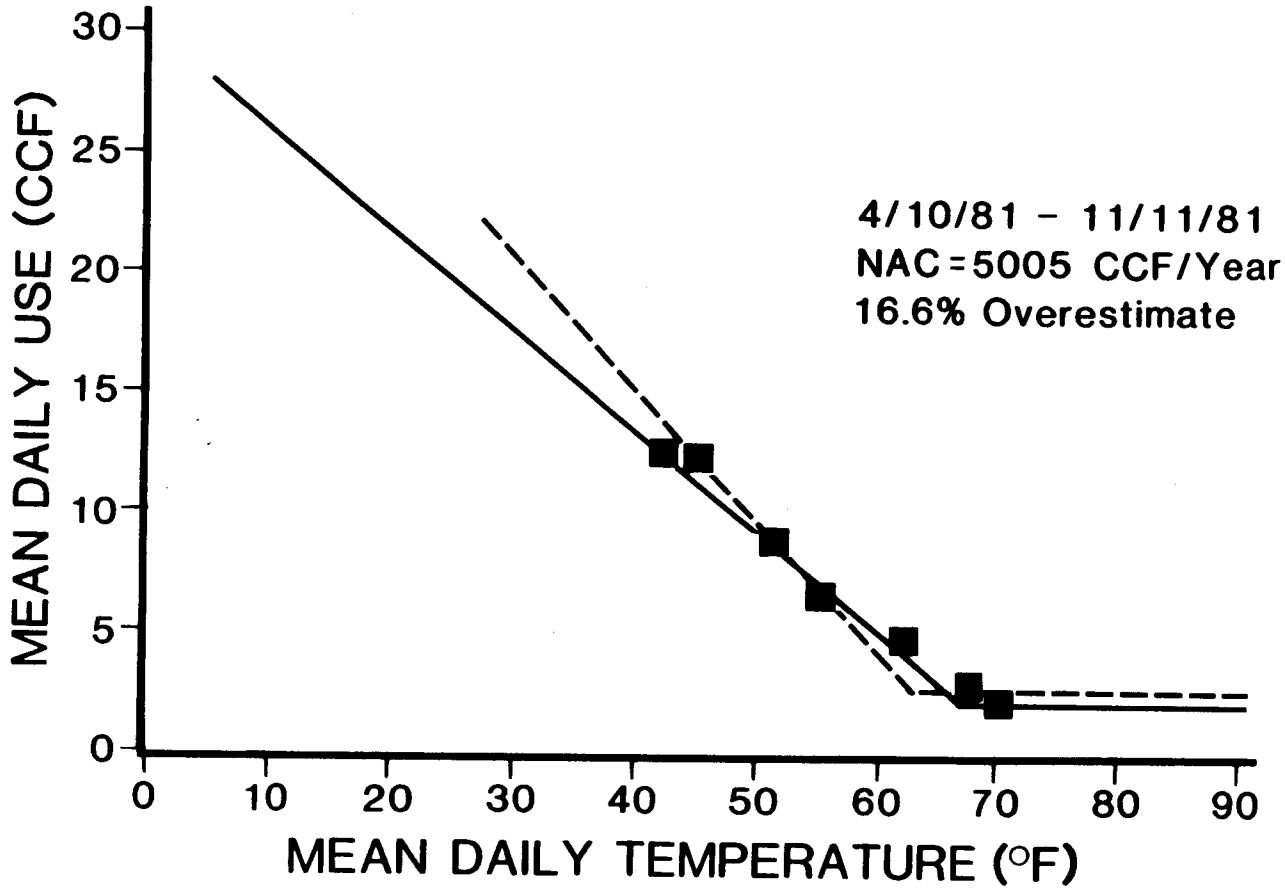
USE VERSUS TEMPERATURE PLOT FOR REPLICATE MISSING SUMMER DATA



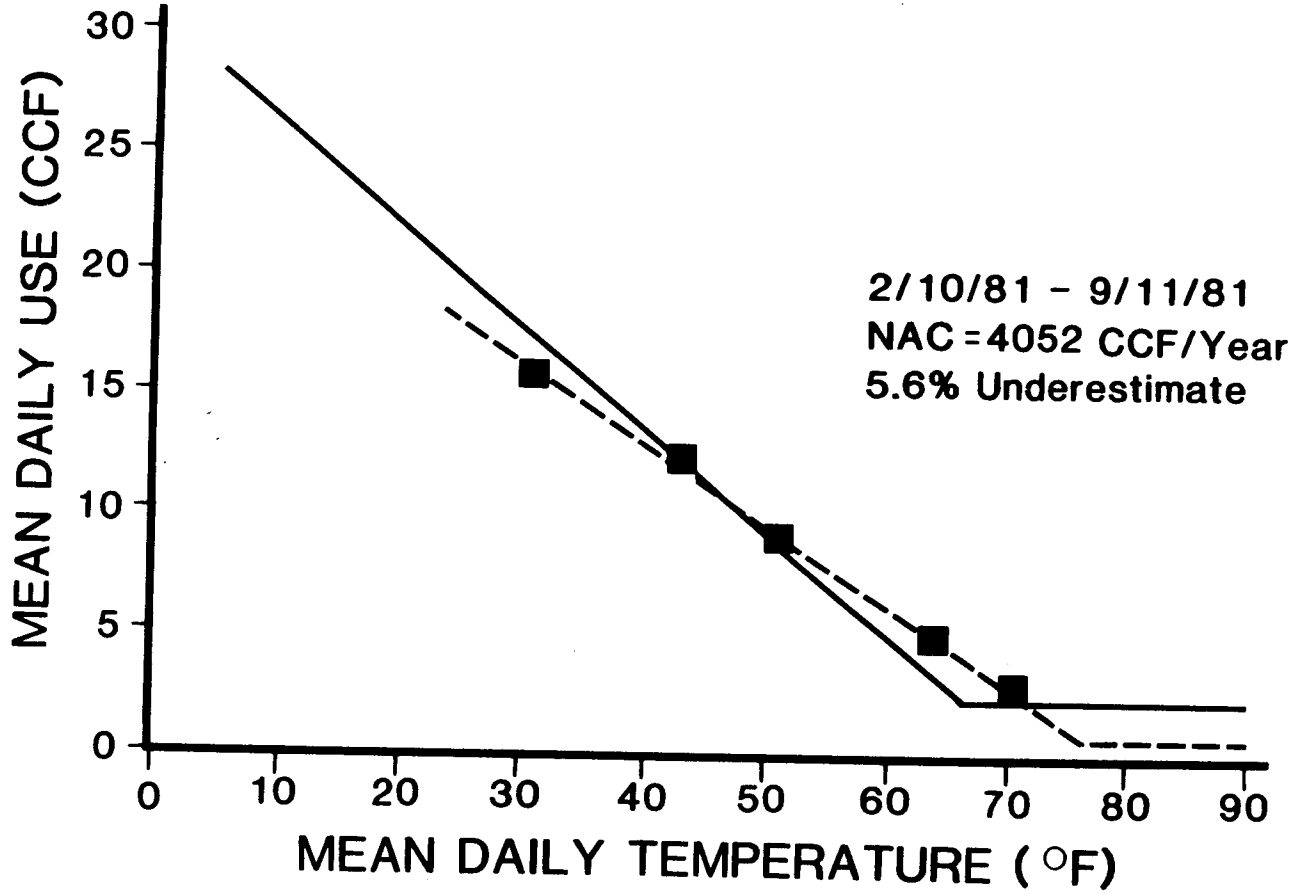
THEORETICAL CHANGE IN APPARENT NAC WHEN SUMMER BASELINE DATA IS ABSENT



USE VERSUS TEMPERATURE PLOT FOR A REPLICATE MISSING WINTER DATA



USE VERSUS TEMPERATURE PLOT FOR A REPLICATE MISSING WINTER DATA

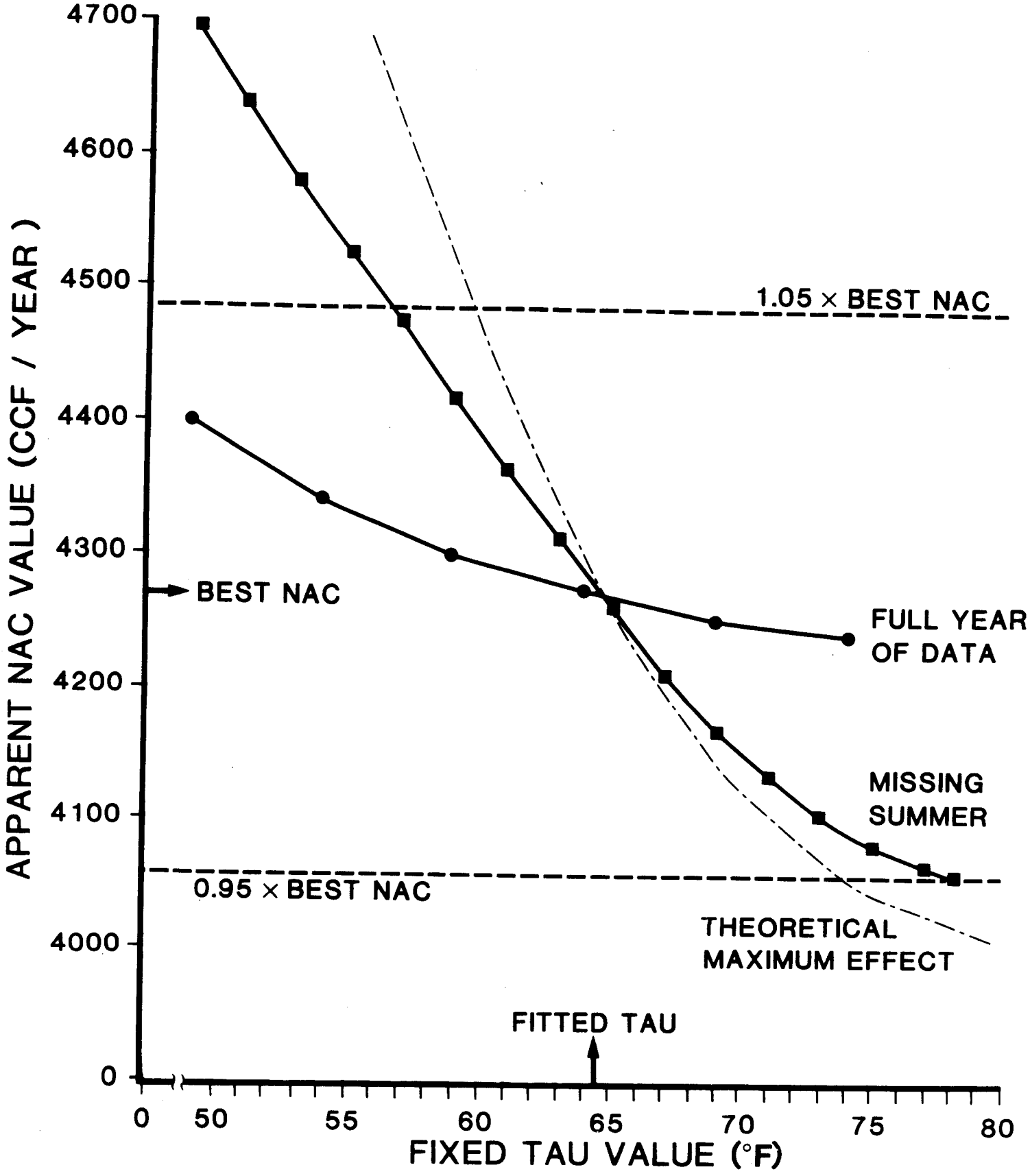


OUTDOOR RESET TEST RESULTS BASED ON INADEQUATE AND ADEQUATE DATA SET LENGTHS

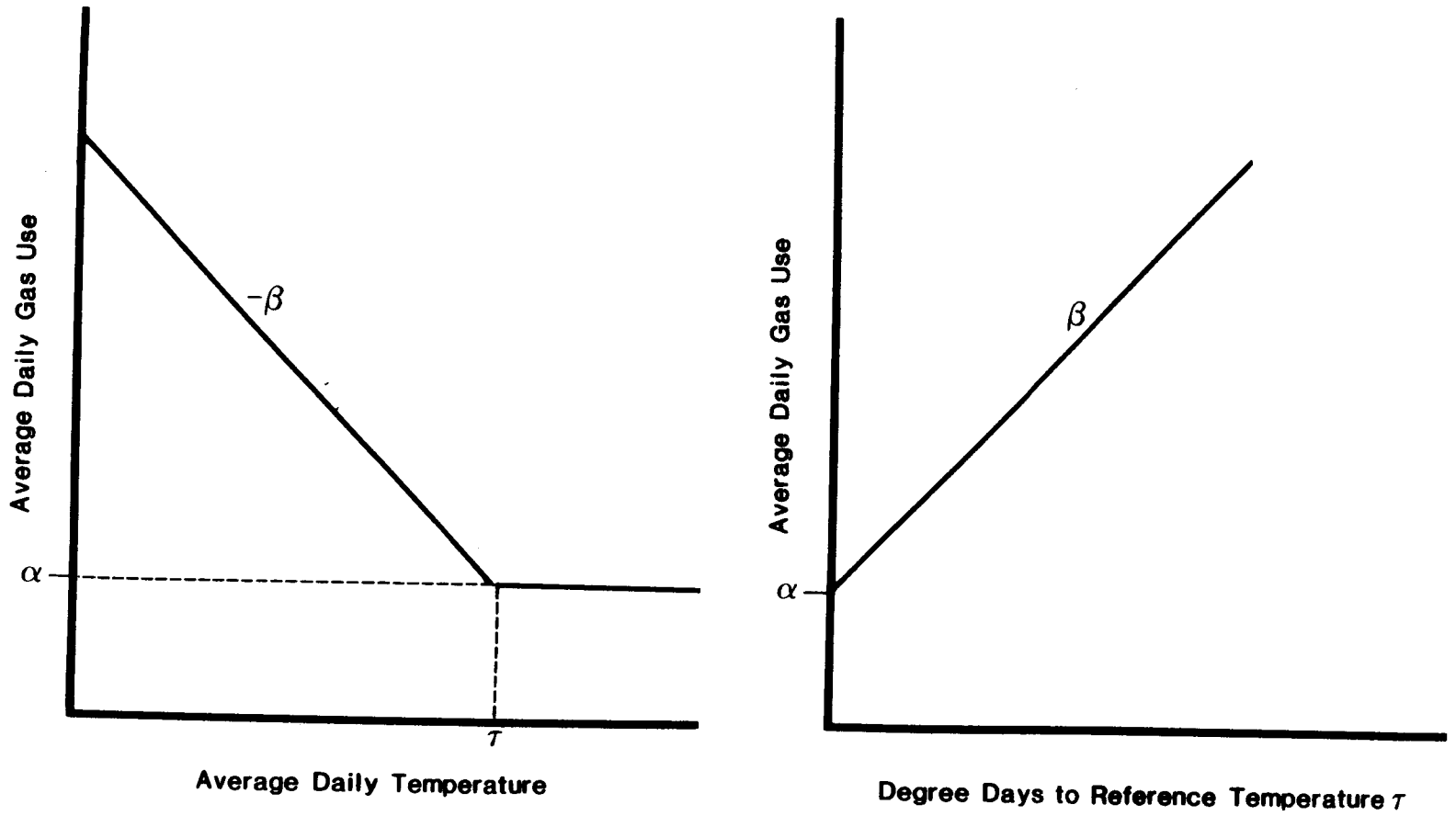
		6 and 7 month		11 and 12 month	
Bldg. ID	Pre NAC	Post NAC	Percent Change	Post NAC	Percent Change
563	9189	6989	-23.9%	8518	-7.3%
111	12089	12303	+1.8	11647	-3.7
133	11083	10242	-7.6	10307	-7.0
1367	13474	9416	-30.1	11276	-16.3
1368	10093	7169	-29.0	8491	-15.9
Means	11186	9224	-17.8%	10048	-10.0%
Std. Dev.	1677	2223	14.2	1491	5.7
t value	(std)	2.602	---	3.452	---
p value	---	p<.10	---	p<.05	---

EFFECT OF VARYING TAU ON DATA WITH AND WITHOUT SUMMER READINGS

(04-366320, 11/79 THROUGH 11/80)



SCOREKEEPING MODEL DEVELOPED BY PRINCETON UNIVERSITY

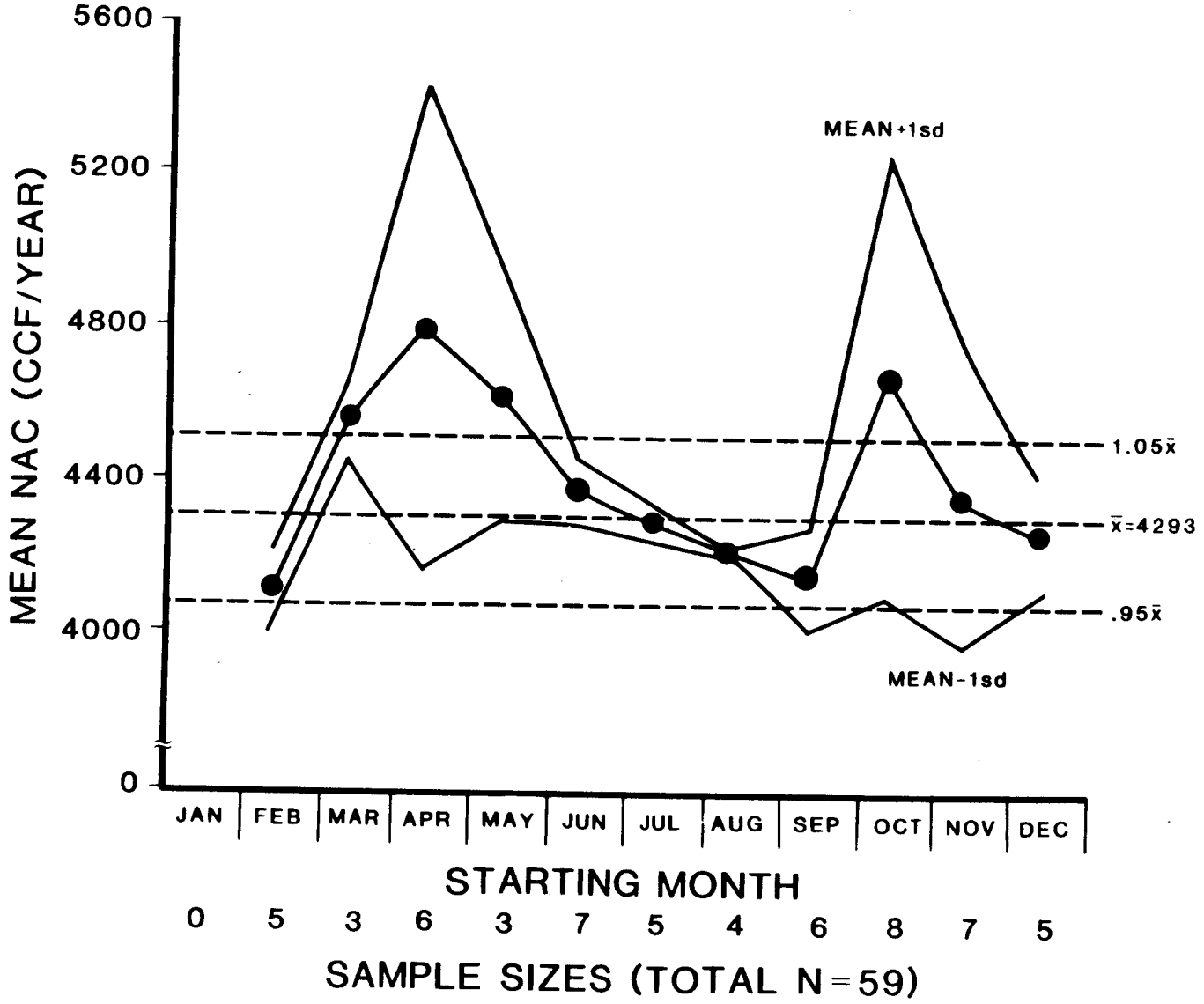


Source: Goldberg, 1982

OVERVIEW OF PRISM DATA CRITERIA

	QUALITY CRITERIA		END USES		
	Input Quality	Output Quality	NAC	ΔNAC	Parameters
most stringent ↕ least stringent	10 - 16 months 1 heating season good summer data no outliers 8 - 24 months some baseload info somewhat clean at least 4 real readings	parameter values and standard errors OK strict R^2/CV criteria relaxed R^2/CV criteria no screening	↕	↕	↕

BIAS AND VARIABILITY OF NAC VERSUS STARTING MONTH FOR REPLICATES WITH SUBSTANDARD AMOUNTS OF DATA (6 - 8 MONTHS)



MIXED - MODEL PRISM RESULTS (12 MONTHS)