Nearest Neighbors Analysis for PRRAP, the Probable Report Rate Analysis Plan

R. L. Obenchain, July 6, 1979

Keywords: Defensible Differences, Performance Ranges

ABSTRACT

The Probable Report Rate Analysis Plan, PRRAP, is an official plan of AT&T Customer Services, Management Analysis. PRRAP is based upon analyses and techniques developed by the Applied Statistics Department, Bell Laboratories, Holmdel. A companion memorandum describes the "standard" features of PRRAP that have already been implemented on the Computer Sharing Services (CSS) system at Denver. This memorandum describes a data analysis procedure, nearest neighbors analysis, that is now available on the Bell Laboratories, Holmdel, computer system.

[NOTE: Because of its "automatic updating" feature, this NN approach was later implemented at CSS …about a year before the official start of Bell System Divestiture in 1982.]

Following a review of PRRAP Models Analysis, the PRRAP Neighbors Analysis concept is introduced and contrasted with the models formulation. The models and neighbors approaches are then shown to complement each other. A numerical example is first used to illustrate the usual case where the two approaches yield similar analyses. Then examples with data errors and extreme facility characteristics illustrate how and why the analyses can differ. Finally, differences in and trade-offs between execution costs and maintenance costs for the two approaches are discussed.

CONTENTS

MEMORANDUM FOR FILE (BELL LABS, TM 79-1711-4)

1. INTRODUCTION

The Probable Report Rate Analysis Plan, PRRAP, allows Bell System managers to compare the customer trouble report rates for individual plant maintenance districts to probable (or benchmark) rates that depend upon environmental and facility characteristics that are beyond the control of local management. A detailed introduction to the definitions and uses of PRRAP is given in the official PRRAP manual [2], and a wealth of technical details about PRRAP procedures is contained in a Bell Laboratories memorandum, [3]. Rather than re-explain here the exact meanings of all of the technical terms used in PRRAP, it is assumed that the reader can consult the above two references, especially the index of the PRRAP manual [2].

This memorandum describes an alternative data analysis procedure, called **PRRAP Neighbors Analysis**, which can be used to "complement" standard **PRRAP Models Analysis** procedures, [2] and [3]. "Standard" PRRAP uses linear regression models to provide "global" statistical predictions of report rates from the values of 13 complexity indicator variables. These variables describe local environmental and facility characteristics and are listed in Table **1.1.** The nearest neighbors analysis procedures described here provide purely "local" comparisons that do not assume that expected report rates are globally linear functions of complexity indicators. These new procedures find the 20 Bell System districts which are most like any given district in terms of all 13 complexity indicators. They then compare that district's 11 report rate components, Table 1.2, with those of its 20 nearest neighbors using PRRAP percentiles and "banding" concepts.

Section 2 illstrates the basic similarities and differences between PRRAP Models and Neighbors Analyses by considering a hypothetical district with characteristics and report rates which are exactly Bell System average. Technical details of the two procedures are given first in Section 3, where the PRRAP models equations are reviewed, and then in Section 4, where the neighbors approach is shown to be compatible with a very general, possibly nonlinear model. Sections 5 and 6 discuss differences between the two approaches in robustness to input errors and to model deficiencies. Section 7 describes possible modifications of PRRAP Neighbors Analysis and also conjectures about their cost implications. Finally, Section 8 provides a brief summary.

2. MODELS AND NEIGHBORS FOR BELL SYSTEM AVERAGE

Before we consider any technical details of how and why PRRAP Models and Neighbors Analyses are either similar or different, let us compare their outputs for a hypothetical district

TABLE 1.1. Complexity Indicators in PRRAP District Data Base

Designation	Designation	
X_1	% RES = % residential stations	Class of Service
X_2	% PBX stations	
X_3	% CTX = % centrex stations	
X_4	RS/L = Residence Stations per Line	
X_5	ACTV = Customer Generated Activity	
X_6	% P&S = % Panel and Step Switching	Central Office & Network
X_7	% XBR = % Crossbar 1 and 5 Switching	
X_8	M/TK = Main Stations Per Trunk	
X_9	CM/L = Conductor Pair Miles Per Line	Outside Plant
X_{10}	% AER = % Aerial Cable	
X_{11}	% UND = % Underground Cable	
X_{12}	PTYL = Party Line Development	
X_{13}	UOF = Urban Operations Factor	

TABLE 1.2. Trouble Report Component Variables

Technical Trouble Variable	Disposition Codes	Designation
Y_1	1+2+3	A123 = Station
Y_2	4	A4 = Outside Plant
Y_3	5+8+0	A580 = Central Office & Network
Y_4	6	A6 = Customer Action
Y_5	7+9	A79 = Not Found
Y_6	$Y_1+Y_2+Y_3+Y_4+Y_5$	ATOT = Total Technical

Administrative Trouble Variable	Designation
Y_7	SUBS = Subsequent Report Percentage
Y_8	EMPL = Employee Report Rate
Y_9	REPT = Repeated Report Percentage
Y_{10}	EXCL = Excluded Report Rate
Y_{11}	MISA = Missed Appointments Percentage

with Bell System average characteristics and report rates. First we will discuss the models output . Then we will introduce the neighbors concept and display the neighbor s output. Finally, we will compare the two outputs.

Figure 2.1 displays the PRRAP Models Analysis f or Bell System average characteristics. The values of the 13 complexity indicator variables for this output are:

```
% RES = 69      % P&S = 30      CM/L  = 2.61
% PBX = 8       % XBR = 48      % AER = 27
% CTX = 5       M/TK  = 15      % UND = 50
RS/L  = 1.68                    PTYL  = 1.05
ACTV  = 52.5                    UOF   = 0
```

These 13 variables are used to estimate the probable (median) report rate for each technical report rate component (Yl to Y6), where "subsequent" reports are not counted. For example , this probable rate for total technical reports (Y6) is 43.9 reports per hundred stations per year and is displayed as the boundary between the "B" and the "C" performance ranges in Figure 2.1.

– Page 7 –

Incidentally, a location which is more trouble prone than Bell System average (say, because % RES < 69 and / or CM/L > 2.61) would be assigned a probable total rate greater than 43. 9. Besides the estimated median rates, PRRAP Models output also displays the estimated 15th , 85th , and 97^{th} percentiles to create A, B, C, D & F performance ranges . Note that the 15% = 37. 3 and 85% = 55.0 for total technical in Figure 2.1 are not equally spaced about the 50% = 43.9. This is the case because PRRAP models, instead of assuming the data are normally or even symmetrically distributed, utilize the empirical distribution of fitted residuals to define percentiles [3]. Specifically, the actual total report rate of 45.9 (not counting "subsequents ") is shown in Figure 2.1 to fall within the *"C"* band at the 60% relative to the probable rate of 43.9 from the PRRAP models. And an actual rate of 45.9 is also shown to fall at the 55% relative to all Bell System districts.

To introduce the nearest neighbors concept, Figure 2.2 is a display of %RES versus GM/L for the 695 districts in the June 1977 PRRAP data base ([3], Section 2). The districts which are inside the circle in Figure 2.2 are thus nearest neighbors to Bell System average on only these two complexity indicators. The distance measure that is actually used in PRRAP Neighbors Analysis replaces circles by hyper-ellipses that account for differences in variance among complexity characteristics.

– Page 9 –

Also, since all 13 complexity indicators are to be used to define nearest neighbors, the circle containing 20 points in Figure 2.2 is replaced by a hyper-ellipse in l3-dimensions that contains 20 points. These 20 districts are described in Tables 2.1 and 2.2. The PRRAP Neighbors

FIGURE 2.1. PRRAP Models Analysis Display for Bell System Average Characteristics

PRRAP MODELS ANALYSIS FOR--- 00000000 BELL SYS

TECHNICAL REPORTS	"A" 01%-15%	"B" 16%-50%	"C" 51%-85%	"D" 86%-97%	"F" 98%-99%	RATE-SUB	GRADE	PRR%/AS%
COMPONENT = 12&3	18.4	21.8	C 25.9	31.2		22.3	C	59/53%
COMPONENT = 4	3.8	5.5	C 8.3	12.3		5.9	C	59/55%
COMPONENT = 5&0	2.5	3.5	C 4.8	7.3		3.7	C	59/58%
COMPONENT = 6	1.1	1.7	C 2.5	3.7		1.9	C	61/58%
COMPONENT = 7&9	8.6	11.6	C 15.4	24.7		12.2	C	58/56%
COMPONENT = TOTL	37.3	43.9	C 55.0	69.9		45.9	C	60/55%

ADMINISTRATIVE REPORTS						RATE	GRADE	AS%
COMPONENT = SUBS	7.2	9.4	C 12.5	17.8		10.2	C	62%
COMPONENT = EMPL	4.1	8.9	C 17.1	27.4		9.1	C	52%
COMPONENT = REPT	9.5	13.3	C 16.9	20.5		13.7	C	56%
COMPONENT = EXCL	5.9	10.8	C 17.1	33.3		10.8	C	51%
COMPONENT = MISA	4.0	6.7	C 10.4	16.4		7.6	C	63%

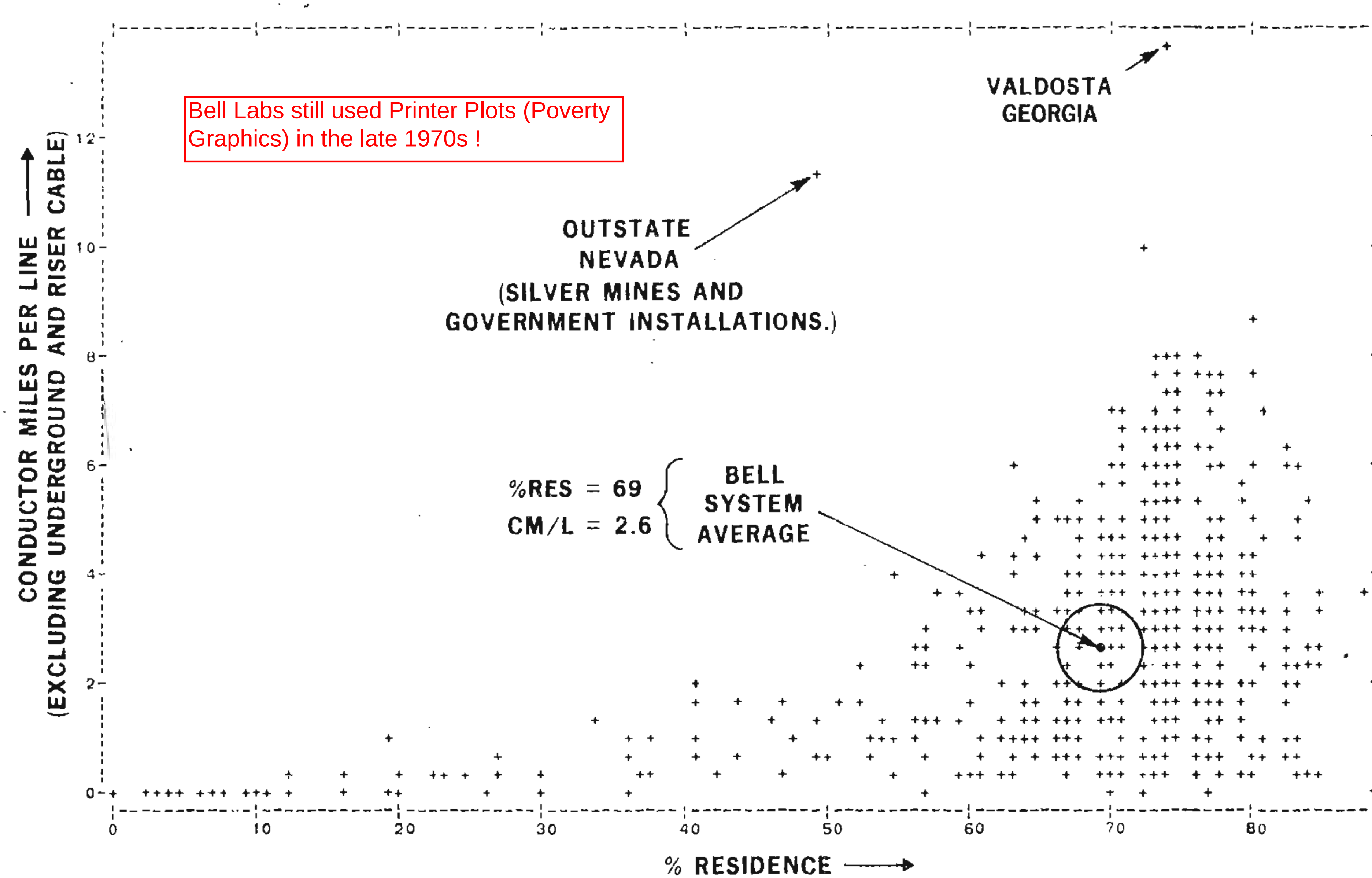

FIGURE 2.2. AN EXAMPLE OF "NEAREST NEIGHBORS" ON %RES AND CM/L

Analysis output display of Figure 2.3 then results from comparing Bell System average trouble report rate components (this time including "subsequents") to the corresponding report components of its 20 nearest neighbors. For example, the third, tenth, and seventeenth order statistics in the sample of size 20 for total technical reports are seen in Figure 2.3 to be 15% = 40.9, 50% = 48 .0, and 85% = 59 . 3, whereas the average actual is 51.1. All of these statistics are slightly larger than the corresponding quantities for PRRAP Models Analysis -- but, essentially, this is only the case because subsequent report s are now included.

There are two more major differences. Neighbors Analysis applies to the administrative trouble variables (Y7 to Yll), whereas all Bell System districts are assigned identical administrative performance ranges in PRRAP Models Analysis. In addition, there are no "F" performance ranges for PRRAP Neighbors Analysis because a sample of size 20 is rather small for reliably estimating an extreme percentile, like the 97 percent point, of a distribution.

TABLE 2.1. The 20 Nearest Neighbors to Bell System Average Characteristics.

Rank	ID Code	Similarity* to Bell System	Company	District
1	02316100	99%	New Eng.	West Rhode Island
2	21541000	99%	Ohio	Dayton North
3	19162700	99%	S. Cent.	Nashville
4	48110003	99%	Cinn.	North
5	27117107	99%	Wisc.	Racine
6	17047130	99%	Sou .	Greenville, SC
7	23144100	99%	Mich.	Lansing
8	48110002	98%	Cinn.	East
9	27117106	98%	Wisc.	Waukesha
10	33300200	98%	S. West	Kansas City
11	31110100	98%	N. West	Cedar Rapids, IA
12	23155500	98%	Mich.	Flint
13	41112400	98%	Pac.	San Mateo
14	23144200	98%	Mich.	Grand Rapids North
15	33300600	97%	S. West	Topeka
16	48110004	97%	Cinn.	West
17	17025240	97%	Sou.	Savanah
18	17047240	97%	Sou.	Charleston
19	31110250	97%	N. West	Souix City West
20	23122100	97%	Mich.	Pontiac

* Similarity percentiles will be explained on page 19.

TABLE 2.2. The 20 Districts with Complexity Indicators most like Bell System Average

THE FOLOWING 20 DISTRICTS ARE YOUR NEAREST NEIGHBORS.

RANK	ID CODE	DIST**2	SIMILARITY TO YOU.
1	02316100	1.62	99%
2	21541000	2.30	99%
3	19162700	2.38	99%
4	48110003	2.41	99%
5	27117107	2.43	99%
6	17047130	2.44	99%
7	23144100	2.47	98%
8	48110002	2.77	98%
9	27117106	2.86	98%
10	33300200	2.91	98%
11	31110100	2.98	98%
12	23155500	3.03	98%
13	41112400	3.08	98%
14	23144200	3.09	97%
15	33300600	3.20	97%
16	48110004	3.24	97%
17	17025240	3.27	97%
18	17047240	3.28	97%
19	31110250	3.30	97%
20	23122100	3.31	97%

OBL	%RES	%PBX	%CTX	RS/L	ACTV	%P&S	%XBR	M/TK	CM/L	%AER	%UND	PTYL	UOF
02316100	67.00	7.00	6.00	1.64	39.51	23.00	54.00	17.00	2.37	33.00	61.00	1.02	0.0
21541000	64.00	8.00	9.00	1.78	53.27	34.00	37.00	11.00	2.01	23.00	64.00	1.01	0.0
19162700	66.00	6.00	10.00	1.79	61.28	30.00	48.00	11.00	2.73	29.00	59.00	1.01	0.0
48110003	71.00	6.00	7.00	1.77	46.19	13.00	54.00	14.00	2.73	42.00	51.00	1.03	0.0
27117107	73.00	7.00	5.00	1.66	44.81	40.00	18.00	22.00	2.69	24.00	57.00	1.03	0.0
17047130	72.00	9.00	1.00	1.76	48.77	37.00	62.00	9.00	4.19	29.00	46.00	1.02	0.0
23144100	67.00	5.00	13.00	1.75	54.41	30.00	35.00	17.00	2.94	27.00	48.00	1.06	0.0
48110002	75.00	4.00	7.00	1.69	50.18	13.00	62.00	13.00	2.55	40.00	56.00	1.03	0.0
27117106	73.00	7.00	2.00	1.78	41.20	3.00	53.00	15.00	3.03	17.00	57.00	1.03	0.0
33300200	73.00	8.00	1.00	1.78	54.48	0.0	49.00	9.00	2.14	21.00	60.00	1.03	0.0
31110100	71.00	7.00	5.00	1.74	53.28	56.00	31.00	20.00	3.26	12.00	47.00	1.06	0.0
23155500	68.00	5.00	7.00	1.78	48.23	57.00	19.00	13.00	3.05	25.00	53.00	1.05	0.0
41112400	68.00	8.00	3.00	1.64	58.06	31.00	43.00	15.00	0.61	26.00	70.00	1.00	0.0
23144200	68.00	7.00	5.00	1.72	50.95	53.00	14.00	9.00	3.43	22.00	50.00	1.07	0.0
33300600	67.00	10.00	4.00	1.74	62.77	24.00	52.00	29.00	3.01	21.00	44.00	1.05	0.0
48110004	65.00	7.00	10.00	1.60	47.29	6.00	61.00	11.00	1.48	30.00	68.00	1.04	0.0
17025240	72.00	11.00	0.0	1.82	48.30	44.00	55.00	19.00	3.36	19.00	50.00	1.01	0.0
17047240	72.00	7.00	4.00	1.87	62.17	42.00	58.00	11.00	3.60	19.00	48.00	1.01	0.0
31110250	73.00	5.00	2.00	1.69	44.67	30.00	41.00	15.00	3.41	9.00	35.00	1.06	0.0
23122100	78.00	3.00	6.00	1.81	53.72	11.00	48.00	11.00	3.49	29.00	49.00	1.02	0.0

COMPARE THESE TO YOUR CHARACTERISTICS OF

	%RES	%PBX	%CTX	RS/L	ACTV	%P&S	%XBR	M/TK	CM/L	%AER	%UND	PTYL	UOF
	69.00	8.00	5.00	1.68	52.47	30.00	48.00	15.00	2.61	27.00	50.00	1.05	0.0

IN THE FOLLOWING PRRAP ANALYSIS, YOUR REPORT RATES WILL BE COMPARED PRIMARILY TO THOSE OF YOUR 20 NEAREST NEIGHBORS-----BUT THE SECOND PERCENTILE PRINTED FOR EACH COMPONENT COMPARES YOU TO THE ENTIRE PRRAP DATA BASE.

FIGURE 2.3. PRRAP Neighbors Analysis Display for Bell System Average Characteristics

```
PRRAP NEIGHBORS ANALYSIS FOR    00000000              BELL SYS

TECHNICAL REPORTS          "A"     "B"     "C"     "D"                     RATE   GRADE   PRR%/AS%
                        01%-15% 16%-50% 51%-85% 86%-99%
-------------------.-------.-------.-------.-------
                   .       .       .  C    .
COMPONENT = 12&3   .    19.5    24.6  C 29.1                               25.0     C      50/53%
                   .       .       .  C    .
-------------------.-------.-------.-------.-------
                   .       .       .  C    .
COMPONENT = 4      .     4.8     6.5  C 10.7                                7.4     C      50/56%
                   .       .       .  C    .
-------------------.-------.-------.-------.-------
                   .       .       .  C    .
COMPONENT = 5&0    .     1.8     3.4  C  5.2                                4.1     C      60/58%
                   .       .       .  C    .
-------------------.-------.-------.-------.-------
                   .       .       .  C    .
COMPONENT = 6      .     1.1     1.5  C  2.4                                2.0     C      75/58%
                   .       .       .  C    .
-------------------.-------.-------.-------.-------
                   .       .       .  C    .
COMPONENT = 7&9    .     8.9    11.1  C 15.0                               12.7     C      55/56%
                   .       .       .  C    .
-------------------.-------.-------.-------.-------
-------------------.-------.-------.-------.-------
                   .       .       .  C    .
COMPONENT = TOTL   .    40.9    48.0  C 59.3                               51.1     C      55/55%
                   .       .       .  C    .
-------------------.-------.-------.-------.-------

ADMINISTRATIVE REPORTS                                                     RATE   GRADE   PRR%/AS%

-------------------.-------.-------.-------.-------
                   .       .  B    .       .
COMPONENT = SUBS   .     8.1  B 10.4    11.8                               10.2     B      45/62%
                   .       .  B    .       .
-------------------.-------.-------.-------.-------
                   .       .       .  C    .
COMPONENT = EMPL   .     4.6     8.4  C 11.7                                9.1     C      55/52%
                   .       .       .  C    .
-------------------.-------.-------.-------.-------
                   .       .       .  C    .
COMPONENT = REPT   .     7.2    11.6  C 14.4                               13.7     C      70/56%
                   .       .       .  C    .
-------------------.-------.-------.-------.-------
                   .       .       .  C    .
COMPONENT = EXCL   .     5.4    10.1  C 16.3                               10.8     C      60/51%
                   .       .       .  C    .
-------------------.-------.-------.-------.-------
                   .       .       .  C    .
COMPONENT = MISA   .     4.3     7.4  C  9.9                                7.6     C      50/63%
                   .       .       .  C    .
-------------------.-------.-------.-------.-------
```

Except for these differences due to handling of administrative reports, especially subsequents, and due to definitions of performance ranges, the outputs of Figures 2.1 and 2.3 are otherwise strikingly similar. Thus the nearest neighbor concept, which is easy to describe and understand, can lead to essentially the same results as do the relatively complicated statistical modeling procedures of "standard" PRRAP analysis. Actually, arguments more important than mere simplicity advantages can be made in favor of neighbors analysis over models analysis. These additional arguments are detailed in Sections 5 and 6 using the terminology and notation established in the next two Sections, 3 and 4.

3. REVIEW OF PRRAP MODELS ANALYSIS

The reader is referred to Section 3 of the companion memorandum [3] for a description and explanation of the multivariate linear regression modeling process that produces PRRAP Models Analysis. This section will repeat only a few of the basic assumptions and formulations for these models. Since the same equation numbers will be used as in [3], missing numbers relate to material that is not being repeated here.

The number of plant maintenance districts (observations) will be denoted by NOBS, and a subscript of i will indicate data from the i-th district $(l \leq i \leq \text{NOBS})$. The number of trouble components (response variables) will be denoted by NVAR, and a subscript j will indicate data on the j-th component $(l \leq j \leq \text{NVAR})$. **Y** will denote the NOBS by NVAR matrix with entry y_{ij} in row i and column j. Thus y_{ij} represents the trouble report rate for district i on component j. The average value of y_{ij} for $1 \leq i \leq \text{NOBS}$ will be denoted by $\bar{y}_{.j}$, and $\bar{\mathbf{y}}'$ will then denote the corresponding row vector (1 by NVAR) of component averages.

The number of complexity indicators (regressor variables) will be denoted by NIND, and a subscript n will indicate data on the n-th indicator $(l \leq n \leq \text{NIND})$. NIND is the same number for each district i and component j. **X** will denote the NOBS by NIND matrix with entries x_{in} and **Z** will denote the NOBS by NIND matrix of "centered" indicators with entries $z_{in} = (x_{in} - \bar{x}_{.n})$. Note that $z_{in} = 0$ implies that district i is average on indicator n, i.e. $x_{in} = \bar{x}_{.n}$. With $\mathbf{z}_i'$ denoting the i-th row of $\mathbf{Z}$, $\mathbf{z}_i' = \mathbf{0}_i'$ would then imply that district i is average on all indicators.

The expectation (first moment) model for PRRAP is then

$$\mathbf{Y} = \mathbf{1}\boldsymbol{\mu}' + \mathbf{Z}\boldsymbol{\beta} + \mathbf{E}\,, \tag{3.1}$$

where **1** is column vector (NOBS by 1) of all ones, $\boldsymbol{\mu}'$ is a row vector (1 by NVAR) of unknown mean values, $\boldsymbol{\beta}$ is a NIND by NVAR matrix of unknown regression coefficients, and **E** is an NOBS by NVAR matrix of unobserved error random variables.

– Page 15 –

Meanwhile, the dispersion (second moment) model for PRRAP states that each row of **Y** has the same variance - covariance matrix, Σ, and all rows are uncorrelated. The dispersion matrix for **Y** can thus be written as the Kronecker product

$$D(\mathbf{Y}) = \boldsymbol{\Sigma} \otimes \mathbf{I} \tag{3. 2}$$

Note that the covariance between y_{ij} and y_{km} is zero unless i = k, in which case this covariance is σ_{jm}, which is the entry in row j and column m of the symmetric matrix Σ. The diagonal elements of Σ are also denoted by $\sigma_{jj} = \sigma_j^2$, so that σ_j also denotes the standard deviation (or standard error) of y_{ij} given **Z** for $1 \le i \le$ NOBS.

These models are used to make "forecasts" as follows. Again letting $\mathbf{z}_i'$ denote the i-th row of **Z**, we note that the row vector of forecast trouble rates, f., at z. are

$$\mathbf{f}_i' = \hat{\boldsymbol{\mu}}' + \mathbf{z}_i' \hat{\boldsymbol{\beta}} \tag{3. 6}$$

where $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\beta}}$ are least squares estimators.

– Page 16 –

The estimated variance-covariance (Dispersion) matrix of these forecasts is

$$D(\mathbf{f}_i') = (1+k_i)\, \boldsymbol{\Sigma} \tag{3.9}$$

where the constant is $k_i = (1 / \text{NOBS}) + \mathbf{z}_i'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z}_i < 1$.

In evaluating the PRRAP models and in calculating the probable report rate percentile (PRR%), we will use ***standardized scores*** defined by

$$\mathbf{s}_{ij} = (y_{ij} - f_{ij}) \,/\, (\hat{\sigma}_j \,/\, \sqrt{1 + k_i}\,) \tag{3.10}$$

which is the difference between the actual report rate and its forecast divided by the forecast standard error. However, instead of assuming that the s-scores have any known distribution, we can store the percentiles (100-Quantiles) of the empirical distribution of s-scores in a matrix **Q** with 99 rows and NVAR columns. The s-scores are first calculated for all districts in which there are no known errors in the data, and this number can be larger than NOBS, the number used to estimate parameters. Then q_{mj} for $1\% \le m \le 99\%$ is defined to be the value below which exactly m percent of these s_{ij} values lie.

These s-quantiles can then be used to define all 99 of the 100-Quantiles of the probable report rate at $\mathbf{z}_i'$ as follows:

$$\hat{p}_{ij}(m) = f_{ij} + q_{mj}\,\hat{\sigma}_j \,/\, \sqrt{1 + k_i}\ , \tag{3.11}$$

which is the estimated m-th quantile of the probable report rate of district i on component j.

4. PRRAP NEIGHBORS ANALYSIS

Suppose now that, instead of a linear model as in Equation (3.1), one has a model for report rates of the form

$$\mathbf{Y} = \mathbf{F} + \mathbf{E}\ , \tag{4.1}$$

where $\mathbf{f}_i'$, the i-th row of the forecast matrix **F**, is some (possibly nonlinear) function of the complexity indicators, $\mathbf{z}'_i$, for the i-th district. To compare the performance of districts i and k on component j, one can calculate the difference in their standardized scores, $s_{ij} - s_{kj}$, from (3.10) and make inferences about the statistical significance of this difference. This difference is now of the general form

$$s_{ij} - s_{kj} = \frac{y_{ij} - f_{ij}}{\sigma_j(i)} - \frac{y_{kj} - f_{kj}}{\sigma_j(k)}\ , \tag{4.2}$$

where $\sigma_j^2(i)$ is the forecast variance at z_{ij}.

Suppose now that the forecasts and the forecast variances are smooth, continuous functions of the complexity indicator vectors, $\mathbf{z_i}$ and $\mathbf{z_k}$, and consider the limit as $\mathbf{z_k}$ approaches $\mathbf{z_i}$. It is intuitively clear that $f_{ij} = f_{kj}$ and $\sigma_j(i) = \sigma_j(k)$ for each j in this limit, and that (4.2) can then be rewritten as

$$s_{ij} - s_{kj} = (y_{ij} - y_{kj}) \times K, \tag{4.2'}$$

because the denominators, $\sigma_j(i)$ and $\sigma_j(k)$, are equal, and the numerator f_{ij} and f_{kj} terms then cancel because they are also equal in (4.2). Of course, the constant factor in (4.2) is then $K = 1/\sigma_j(i) = 1/\sigma_j(k)$. In other words, without even specifying which nonlinear model is appropriate (let alone estimating its parameters), one knows that all differences in scores will become directly proportional to differences in observed report rates in the limit as two districts have identical $\mathbf{z}'$ vectors of complexity characteristics.

The basis of PRRAP Models Analysis is that using “scored” report rates makes all districts directly comparable. In sharp contrast, PRRAP Neighbors Analysis merely compares the actual report rates of those districts which have most nearly equal complexity characteristics.

To accomplish this, PRRAP Neighbors Analysis performs calculations to determine the 19 “nearest neighbors” of any given district. The PRRAP data base contains all of the information needed to compute the complexity indicator vector, $\mathbf{z}'$, for each of the NOBS = 695 plant maintenance districts in the Bell System. Given the complexity indicator vector for the i-th district, $\mathbf{z}_i'$, the necessary computational task is to compute its "distance" from the 694 other $\mathbf{z}_k'$ vectors with k ≠ i. One possibility, analogous to the calculations in (3.9), would be to calculate squared distances of the form: $(\mathbf{z}_i' - \mathbf{z}_k')(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{z}_i - \mathbf{z}_k)$. To avoid having to invert the $\mathbf{Z}'\mathbf{Z}$ matrix, the alternative outlined below ignores the off-diagonal elements of this matrix. The necessary calculation is then of the form:

$$d_i^2(\mathbf{z}_k') = \sum_{n=1}^{NIND} (z_{in} - z_{kn})^2 / SS(n) \qquad (4.3)$$

where $SS(n) = \sum_{k=1}^{NOBS} z_{kn}^2$ is proportional to the sample variance of indicator n.

[NOTE: The official implementation of PRRAP NN at CSS actually did the matrix inversion referenced above and in (7.1) below. Thus, rather than the shortcut of (4.3), a Mahalanobis-like distance was ultimately used in the measurement plan.]

The d^2 values from (4.3) are then formed into a column vector and joined with the **Y** and **Z** matrices to form a NOBS by (NVAR + NIND + 1) matrix of the form (**Y** | **Z** | $\mathbf{d^2}$), and the row vectors of this composite matrix are sorted into increasing order on the last column. Since $d^2()$ is always zero for k = i, the row for the initially given (i-th) district usually ends up at the very top of this sorted, composite matrix. The top 20 sorted rows of **Y** contain the actual report rates for these 20 "nearest neighbors" to the i-th district, and the top 20 sorted rows of **Z** contain the $\mathbf{z}_k$ vectors most like $\mathbf{z}_i$. Only these top 20 rows are needed to perform PRRAP Neighbors analysis for the i-th plant maintenance district.

– Page 19 –

Since $\mathbf{z}' = \mathbf{0}'$ corresponds (by definition) to Bell System average complexity characteristics, the 100-quantiles of the empirical distribution of $d^2(\mathbf{0}')$ are stored in the PRRAP data base. As explained next, these quantiles are used to define "complexity likeness percentiles" for all nearest neighbor districts relative to any given district $\mathbf{z}'$vector.

If a given $\mathbf{z}'$ vector represents either extreme characteristics or else an unusual combination of complexity characteristics, then the 20 smallest resulting $d^2(\mathbf{z}')$values will still be rather large compared to the observed distribution of $d^2(\mathbf{0}')$ values. The output for PRRAP neighbors analysis includes a "smilarity to you" percentile [see Table 2.2] which is defined to be the percentage of $d^2(\mathbf{0}')$ values which are ***larger than*** an observed $d^2(\boldsymbol{z}')$ value. These similarity percentiles will all be between 95% and 99% when the nearest neighbors of $\boldsymbol{z}'$are indeed as close to $\boldsymbol{z}'$ as is common, but they can all be below 50% if $\boldsymbol{z}'$ is extreme or unusual.

Meanwhile, the ***performance range*** table for PRRAP Neighbors Analysis is computed directly from the top 20 rows of the **Y** matrix after the above sorting. Suppose that y_j^* for 1 ≤ j ≤ NVAR denotes the j-th trouble rate component for the district with complexity characteristics **z***, and let m be the integer between 0 and 20 inclusive such that, of the first 20 entries in the j-th column of

Y, exactly m are less than or equal to y_j^*. Then, in analogy with (3.11), the neighbor PRR% of y_j^* is five times m.

– Page 20 –

Finally, only the top 20 rows of the j-th column of **Y** are sorted …so that $y_{1j} \leq y_{2j} \leq \ldots \leq y_{20j}$ become the observed "order statistics" of the j-th component of nearest neighbor report rates. In particular, y_{10j} is the nearest neighbor PRR (or "B"/"C" boundary), y_{3j} is the "A"/"B" boundary, and y_{17j} is the "C" / "D" boundary. Due to the relatively small sample of only 20 neighbors, no "F" band is estimated for display in the NN performance range table.

5. ROBUSTNESS TO INPUT ERRORS

Table 5.1 illustrates the types of warnings that are issued by PRRAP when an "unusual" combination of complexity indicator values are input. In this case, attention is drawn to the large value of eMIL = 13.68, which was an obvious outlier in Figure 2.2. The PRRAP models analysis display of Figure 5.1 is thus possibly misleading.

The nearest neighbor calculations of Table 5.2 show that 19 of Valdosta's nearest neighbors have CM/L < 8. There are 8 districts in the June 1977 PRRAP data base with CM/L> 8, but only one of these is among Valdosta's nearest neighbors. This suggests that CM/L = 13.68 is an input error and that Valdosta's correct CM/L is probably much less than 13.68. (The "corrected" value entered in April 1978 is CM/L = 7.45.) Thus, it is only due to an input error on CM/L that Valdosta 's nearest neighbors seem rather distant (similarity only ~ 27%.)

– Page 25 –

In fact, if the input error were corrected, sixteen of Valdosta's true nearest neighbors would still be among the 20 districts identified with the incorrect value of CMIL = 13.68. And this means that the PRRAP Neighbors Analysis display of Figure 5.2 is already the essentially correct one even though CMIL = 13.68 is an input error. Note, in particular, that the neighbor PRR%'s of Figure 5.2 are either much higher or much lower than the misleading PRRAP Model PRR%'s of Figure 5.1.

PRRAP Models Analysis is clearly not robust to input errors on complexity indicator variables. PRRAP Neighbors Analysis is robust to these errors if and only if all nearest neighbors become distant (low similarity.) Any input error which moves a district away from its "true" neighbors but close to "false" neighbors will probably not be detected by either PRRAP Models or PRRAP Neighbors Analyses.

Also, neither PRRAP Models nor Neighbors Analysis is robust to input errors in report rate components.

6. ROBUSTNESS TO MODEL DEFICIENCIES

It has been stressed that PRRAP Neighbors Analysis is compatible with any linear or smooth nonlinear model. PRRAP Models Analysis is not only based upon a globally linear model but

TABLE).1. Warnings Issued by PRRAP that Indicate Possible Input Errors.

```
PRRAP MODELS FOR---   17025220   VALDESTA   OUTSTATE

CHARACTERISTICS =

 PERCENT RESIDENCE OF      74.00

 PERCENT PBX OF       7.00

 PERCENT CTX OF       0.0

 RESIDENCE STATIONS PER LINE OF     1.7805

 ACTIVITY INDEX OF     41.53

 PERCENT PANEL AND STEP SWITCHING OF     40.00

 PERCENT CROSS BAR SWITCHING OF     59.00

 MAIN STATIONS PER TRUNK OF     20.00

 CONDUCTOR MILES PER MAIN LINE OF     13.680
THIS VALUE IS OUTSIDE OF THE EXPECTED RANGE FROM    0.0   TO   10.000.

 PERCENT AERIAL (EXCEPT RISER) CABLE OF      57.00

 PERCENT UNDERGROUND AND RISER CABLE OF      6.00

 PARTY LINE RATIO OF     1.05

 URBAN OPERATIONS FACTOR OF      0.0

 WARNING------THIS FACILITIES MIX REPRESENTS AN EXTREMELY UNUSUAL COMBINATION OF CHARACTERISTICS RELATIVE TO AVAILABLE
 BELL SYSTEM DATA. (CHECK FOR ERRORS IN YOUR FACILITIES INPUTS.)  EXECUTION WILL CONTINUE, BUT A POSSIBLY UNWARRANTED
 EXTRAPOLATION OF THE PRRAP MODELS MAY RESULT.
```

Range Warning (See [3], Table 2.9, page 31.)

Leverage Warning (See [3], page 41.)

FIGURE 5.1. A Misleading PRRAP Models Display
due to an Error on CM/L.
(Compare This Display to Figure 5.2.)

PRRAP MODELS ANALYSIS FOR--- 17025220 VALDOSTA OUTSTATE

TECHNICAL REPORTS	"A" 01%-15%	"B" 16%-50%	"C" 51%-85%	"D" 86%-97%	"F" 98%-99%	RATE-SUB	GRADE	PRR%/AS%
COMPONENT = 12&3	27.0	B 30.7	35.1	40.8		27.1	B	16/86%
COMPONENT = 4	A 11.2	13.1	16.2	20.4		10.4	A	7/89%
COMPONENT = 58&0	0.8	1.9	C 3.3	6.0		2.2	C	61/18%
COMPONENT = 6	0.5	B 1.2	2.0	3.3		1.2	B	55/25%
COMPONENT = 7&9	9.0	12.3	C 16.4	26.5		13.7	C	65/58%
COMPONENT = TOTL	51.7	B 58.9	70.8	87.0		54.5	B	27/81%

ADMINISTRATIVE REPORTS						RATE	GRADE	AS%
COMPONENT = SUBS	7.2	9.4	C 12.5	17.8		9.9	C	57%
COMPONENT = EMPL	4.1	8.9	17.1	D 27.4		19.0	D	89%
COMPONENT = REPT	9.5	B 13.3	16.9	20.5		11.9	B	36%
COMPONENT = EXCL	5.9	10.8	C 17.1	33.3		11.5	C	56%
COMPONENT = MISA	4.0	B 6.7	10.4	16.4		6.4	B	46%

TABLE 5.2. Nearest Neighbor Statistics for Valdosta Outstate with Input Error on CM/L.

THE FOLOWING 20 DISTRICTS ARE YOUR NEAREST NEIGHBORS.

RANK	ID CODE	DIST**2	SIMILARITY TO YOU.
1	16003130	14.29	27%
2	17036150	15.52	24%
3	19111500	16.40	23%
4	17025130	16.47	23%
5	19111600	18.04	19%
6	17025140	18.67	19%
7	19123800	18.78	19%
8	17025110	18.88	19%
9	19162500	19.04	19%
10	19112500	19.60	18%
11	19161300	20.03	17%
12	19153100	20.23	17%
13	17047230	20.23	17%
14	19161400	20.29	17%
15	47800000	20.45	17%
16	19163800	20.49	17%
17	19162600	20.57	17%
18	02518100	20.69	17%
19	19152300	21.08	17%
20	16003430	21.66	16%

All nearest neighbors are distant from the complexity indicator combination that was input.

OBL	%RES	%PBX	%CTX	RS/L	ACTV	%P&S	%XBR	M/TK	CM/L	%AER	%UND	PTYL	UOF
16003130	73.00	7.00	0.0	1.57	52.17	28.00	69.00	16.00	7.92	35.00	10.00	1.09	0.0
17036150	77.00	6.00	2.00	1.69	39.27	68.00	31.00	29.00	7.33	41.00	11.00	1.04	0.0
19111500	76.00	7.00	0.0	1.56	40.24	30.00	70.00	28.00	7.88	30.00	16.00	1.00	0.0
17025130	80.00	4.00	0.0	1.74	34.66	21.00	79.00	34.00	8.81	27.00	10.00	1.12	0.0
19111600	78.00	6.00	0.0	1.55	38.97	34.00	66.00	25.00	7.19	30.00	16.00	1.01	0.0
17025140	75.00	7.00	0.0	1.72	43.39	71.00	19.00	21.00	6.68	38.00	18.00	1.04	0.0
19123800	75.00	5.00	4.00	1.45	48.22	59.00	40.00	21.00	7.38	75.00	7.00	1.11	0.0
17025110	71.00	6.00	5.00	1.75	55.89	57.00	39.00	23.00	6.87	31.00	24.00	1.05	0.0
19162500	81.00	3.00	0.0	1.58	53.27	65.00	35.00	30.00	6.95	40.00	12.00	1.06	0.0
19112500	71.00	8.00	3.00	1.81	52.33	39.00	60.00	19.00	6.73	25.00	21.00	1.02	0.0
19161300	76.00	7.00	0.0	1.48	41.91	46.00	50.00	42.00	7.60	41.00	12.00	1.08	0.0
19153100	78.00	4.00	0.0	1.74	44.09	63.00	33.00	26.00	6.74	29.00	10.00	1.10	0.0
17047230	75.00	9.00	0.0	1.71	37.65	63.00	36.00	19.00	6.81	24.00	19.00	1.05	0.0
19161400	76.00	7.00	0.0	1.48	41.91	46.00	50.00	42.00	7.53	41.00	13.00	1.08	0.0
47800000	74.00	7.00	3.00	1.65	39.78	80.00	17.00	26.00	6.62	68.00	26.00	1.03	0.0
19163800	77.00	6.00	0.0	1.50	42.32	47.00	42.00	19.00	7.30	22.00	8.00	1.08	0.0
19162600	73.00	7.00	2.00	1.51	45.58	43.00	45.00	27.00	5.74	51.00	12.00	1.04	0.0
02518100	73.00	7.00	2.00	1.58	39.99	66.00	33.00	18.00	6.32	61.00	12.00	1.14	0.0
19152300	74.00	6.00	3.00	1.61	49.76	53.00	46.00	30.00	7.45	43.00	13.00	1.19	0.0
16003430	74.00	7.00	2.00	1.55	50.73	57.00	39.00	25.00	5.08	38.00	27.00	1.07	0.0

COMPARE THESE TO YOUR CHARACTERISTICS OF

CM/L is nearly twice that of all nearest neighbors.

	%RES	%PBX	%CTX	RS/L	ACTV	%P&S	%XBR	M/TK	CM/L	%AER	%UND	PTYL	UOF
	74.00	7.00	0.0	1.78	41.53	40.00	59.00	20.00	13.68	57.00	6.00	1.05	0.0

IN THE FOLLOWING PRRAP ANALYSIS, YOUR REPORT RATES WILL BE COMPARED PRIMARILY TO THOSE OF YOUR 20 NEAREST NEIGHBORS-----BUT THE SECOND PERCENTILE PRINTED FOR EACH COMPONENT COMPARES YOU TO THE ENTIRE PRRAP DATA BASE.

FIGURE 5.2. The PRRAP Neighbors Analysis Display can Give Reasonable Evaluations Even When Input Errors Occur. (Compare This Display to Figure 5.1.)

PRRAP NEIGHBORS ANALYSIS FOR 17025220 VALDOSTA OUTSTATE

TECHNICAL REPORTS	"A" 01%-15%	"B" 16%-50%	"C" 51%-85%	"D" 86%-99%	RATE	GRADE	PRR%/AS%
COMPONENT = 12&3	20.5	26.0	C 31.9		29.7	C	75/82%
COMPONENT = 4	7.3	B 13.1	16.6		12.6	B	45/86%
COMPONENT = 58&0	A 2.9	3.2	3.8		2.3	A	5/18%
COMPONENT = 6	A 1.3	1.8	2.4		1.2	A	5/25%
COMPONENT = 7&9	10.0	12.8	C 19.2		14.1	C	55/67%
COMPONENT = TOTL	47.3	57.2	C 73.9		60.0	C	60/78%

ADMINISTRATIVE REPORTS	"A" 01%-15%	"B" 16%-50%	"C" 51%-85%	"D" 86%-99%	RATE	GRADE	PRR%/AS%
COMPONENT = SUBS	6.0	9.3	C 12.3		9.9	C	60/57%
COMPONENT = EMPL	5.7	8.6	18.4	D	19.0	D	85/89%
COMPONENT = REPT	10.5	B 16.5	18.7		11.9	B	25/36%
COMPONENT = EXCL	4.8	6.3	C 13.4		11.5	C	80/56%
COMPONENT = MISA	3.6	B 6.4	12.7		6.4	B	45/46%

also upon the specific estimated parameters of that model. Thus the basic assumptions underlying Neighbors Analysis are not violated unless those underlying Models Analysis are also inappropriate.

– Page 26 –

In fact, if a set of data yields contradictory PRRAP Neighbors and Models Analyses, the Models Analysis is almost surely "wrong". However, the Neighbors Analysis could still be reasonably "correct," as in the previous section of this memorandum on input errors.

Figures 6.1 and 6.2 are the Models and Neighbors Analyses for the state of Nevada (except Reno), which (besides Valdosta) is the other obvious outlier in Figure 2.2. The Nevada inputs are apparently correct; due to large numbers of silver mines and government installations, Nevada has a truly unusual combination of low % RES = ~9 and high CM/L = 11.23. Extrapolating the PRRAP Models to provide the forecasts of Figure 6.1 produced extremely high probable rates. On the other hand, the PRRAP Neighbors Analysis display of Figure 6.2 gives a realistic evaluation of Nevada, especially relative to not-found (7&9) and total troubles.

Again, PRRAP Models Analyses are not robust. Besides yielding unreasonable results upon extrapolation, there may be regions in 13-dimensional complexity indicator space where report rates systematically deviate from globally linear models. Neighbor comparisons are robust to model deficiencies and thus will almost always yield reasonable results.

– Page 29 –

7. GENERALIZATIONS AND COST IMPLICATIONS

The PRRAP Neighbors Analysis software presently available on the Bell Laboratories, Holmdel, computer is about three times as expensive (in terms of execution time) as is PRRAP Models Analysis. It will be argued that some generalizations of the simple formulation of Section 4 might further increase the execution time of neighbors software. On the other hand, costs related to errors in updating of inputs and reprogramming software to detect such errors may be avoided by using neighbors to complement models. The generalizations of neighbors analysis that will be discussed give users options to change the distance formula, (4.3), and/or the number of neighbors to be sought. The generalized squared distances are

$$d_i^2(\mathbf{z}') = (\mathbf{z}_i' - \mathbf{z}')\mathbf{W}S^{-1}\mathbf{W}(\mathbf{z}_i - \mathbf{z}'). \qquad (7.1)$$

where $\mathbf{S}$ is a NIND by NIND positive definite "scaling" matrix and $\mathbf{W}$ is a diagonal matrix of nonnegative "weights." Equation (4.3) is the special case where $\mathbf{S}$ is proportional to the diagonal matrix of the variances of the complexity indicators and $\mathbf{W}$ is the identity matrix.

Users could be given options to set a proper subset of the NIND weights, $\mathbf{W}$, to zero or even to input any nonnegative set with a positive sum. And $\mathbf{S}$ could be chosen to be the entire variance-covariance matrix of complexity indicators instead of only its diagonal. Since the quadratic form of (7.1) yields contours of constant distance that are hyper-ellipsoids centered at $\mathbf{z}'$, this latter

FIGURE 6.1. Example of a Drastic Extrapolation of the PRRAP Models

Note the extremely large predicted station (12 & 3), not-found (7 & 9) and total trouble rates.

PRRAP MODELS ANALYSIS FOR--- 41312100 STATE NEVADA B

TECHNICAL REPORTS	"A" 01%-15%	"B" 16%-50%	"C" 51%-85%	"D" 86%-97%	"F" 98%-99%	RATE-SUB	GRADE	PRR%/AS%
COMPONENT = 12&3	A 28.2	31.8	36.2	41.8		17.0	A	1/ 6%
COMPONENT = 4	A 10.5	12.4	15.4	19.5		6.3	A	1/60%
COMPONENT = 58&0	4.0	5.0	6.5	D 9.1		7.0	D	90/96%
COMPONENT = 6	1.6	B 2.1	3.1	4.3		2.1	B	48/67%
COMPONENT = 7&9	A 14.2	17.3	21.4	31.2		14.1	A	15/74%
COMPONENT = TOTL	A 61.7	68.7	80.3	96.1		46.6	A	1/57%

ADMINISTRATIVE REPORTS	"A" 01%-15%	"B" 16%-50%	"C" 51%-85%	"D" 86%-97%	"F" 98%-99%	RATE	GRADE	AS%
COMPONENT = SUBS	A 7.2	9.4	12.5	17.8		4.2	A	2%
COMPONENT = EMPL	A 4.1	8.9	17.1	27.4		3.0	A	9%
COMPONENT = REPT	9.5	13.3	16.9	D 20.5		19.0	D	94%
COMPONENT = EXCL	A 5.9	10.8	17.1	33.3		4.1	A	9%
COMPONENT = MISA	4.0	6.7	10.4	D 16.4		11.4	D	91%

FIGURE 6.2. The PRRAP Neighbors Analysis Display can Give Reasonable Evaluations Even When all Nearest Neighbors are Distant

PRRAP NEIGHBORS ANALYSIS FOR 41312100 STATE NEVADA B

TECHNICAL REPORTS	"A" 01%-15%	"B" 16%-50%	"C" 51%-85%	"D" 86%-99%	RATE	GRADE	PRR%/AS%
COMPONENT = 12&3	A 20.9	24.4	27.1		17.8	A	5/ 3%
COMPONENT = 4	6.7	B 10.2	13.5		6.9	B	20/51%
COMPONENT = 58&0	2.9	4.1	C 7.3		7.3	C	85/93%
COMPONENT = 6	1.6	B 2.8	3.7		2.2	B	30/63%
COMPONENT = 7&9	9.1	12.4	C 18.5		14.4	C	65/69%
COMPONENT = TOTL	48.0	B 52.5	64.6		48.6	B	25/46%

ADMINISTRATIVE REPORTS					RATE	GRADE	PRR%/AS%
COMPONENT = SUBS	A 4.5	7.4	9.6		4.2	A	10/ 2%
COMPONENT = EMPL	1.0	B 8.4	13.3		3.0	B	20/ 9%
COMPONENT = REPT	9.2	13.8	C 19.0		19.0	C	85/94%
COMPONENT = EXCL	0.5	3.7	C 9.8		4.1	C	60/ 9%
COMPONENT = MISA	3.1	5.7	10.5	D	11.4	D	95/91%

option allows the user to orient the principal axes of the hyper-ellipse so as to account for the correlations among complexity indicators as well as for differences in variance.

– Page 30 –

The above options would almost double computation time unless a "similarity to you" percentile is also made optional. If this percentile is requested, $d_i^2(\mathbf{0}')$ for $1 \leq i \leq NOBS = 695$ must be recomputed using the optional weights and scaling. Then, rather than first sorting all of the $d_i^2(\mathbf{0}')$ values, it is cheaper to simply count how many of them exceed each of the twenty smallest $d_i^2(\mathbf{z}')$ values. Specifically, the similarity percentile of $d_i^2(\mathbf{z}')$ is the number of $d_i^2(\mathbf{0}')$ values that exceed $d_i^2(\mathbf{z}')$ divided by NOBS/100 = 6.95.

To clarify the above remark about the relative costs of sorting and counting, let us now reconsider the problem of reordering the rows of the **Y** and **Z** matrices so that the first 20 rows correspond to the 20 smallest squared distances, $d_i^2(\mathbf{z}')$, of either (4.3) or (7.1). A simple but relatively expensive way to do this would be to form a partitioned matrix with NOBS = 695 rows and (NIND + 12) columns as suggested earlier and sort this matrix using STATLIB[l] subroutine STSORT. It would be more economical to calculate the 20-th order statistic of the $d_i^2(\mathbf{z}')$ values

– Page 31 –

using STATLIB[l] subroutine STFNOB with function argument ORD. One then cycles through the rows of the composite ($\mathbf{Y} \mid \mathbf{Z} \mid \mathbf{d}^2$) matrix looking for rows with $d_i^2(\mathbf{z}')$ values ≤ this threshold. Only these rows (20 or more, including ties) would be transferred to a separate matrix. Finally, the rows of this sub-matrix would be sorted on each column of **Y**.

These simplified neighbors calculations would still be more complicated to program and at least three times as expensive to run as the models routines. But they would provide Bell System managers with analyses which, without programming changes, would always be as up-to-date as the most recent PRRAP data base.

Specifically, the PRRAP Models routines would continue to compare a mangers' new data to the June 1977 PRRAP data base, allowing him/her to see trends for his/her district relative to a set of fixed, historical benchmarks. But PRRAP Neighbors routines would allow a manager to compare his/her current report rates to those of the other Bell System districts which, in terms of their current complexity characteristics, are now most like his/her OWN ever evolving complexity.

8. SUMMARY

The regression models routines in PRRAP provide "global" statistical predictions of report rates using multiple linear regression techniques. In quantifying magnitudes of deviations from the benchmarks in terms of PRR%'s, these models do not assume that the deviations follow a symmetrical, normal distribution. The PRRAP models routines provide statistically validated analyses that can be submitted to public utility commissions. And the use of these analysis procedures can be sold or leased to telephone companies outside of the Bell System.

However, models analyses are not "robust" to input data errors or to extrapolations of complexity beyond regions where data are available.

The neighbors analysis routines in PRRAP provide "local" comparisons which do not assume that expected report rates are globally linear functions of complexity indicators. These routines list the 20 Bell System districts which are most like any given district in terms of all 13 complexity indicators. They then compare that district's report rates with the report rates of its 20 nearest neighbors using the PRR% and "banding" concepts. These neighbor analyses display *extensive* Bell System proprietary information which excludes commercial use outside the Bell System and which demands careful explanation when presented to public utility commissions. But these analyses complement model analyses by being "robust" to not only regression model deficiencies but also to input data errors and extrapolations.

REFERENCES

[l] Brelsford, W. M. and Relles, D. A. (1978). **STATLIB:** A Statistical Computing Library.- Users Manual. Bell Laboratories Applied Statistics Department, West Long Branch, NJ.

[2] Gorry, P. E., Heathcock, J. R., and Saalfrank, J. C. (1978). Probable Report Rate Analysis Plan. AT&T Customer Services Department, Management Analysis Section.

[3] Obenchain, R, L. (1979). Data Analysis and-Statistical Modeling for PRRAP, the Probable Report Rate Analysis Plan. Bell Laboratories TM-79-1711-02.