POOLING GENERATING UNIT DATA FOR IMPROVED ESTIMATES OF PERFORMANCE INDICES

FINAL REPORT

of the

TASK FORCE ON GENERATING UNIT DATA POOLING

APPLICATION OF PROBABILITY METHODS SUBCOMMITTEE

POWER SYSTEM ENGINEERING COMMITTEE

IEEE POWER ENGINEERING SOCIETY

Task Force on Generating Unit Data Pooling: A.M. Breipohl, Chairman, P.F. Albrecht, R.N. Allan, S. Asgarpoor, M.P. Bhavaraju, R. Billinton, G.M. Curley, C.R. Heising, M. Mazumdar, M.F. McCoy, R.J. Niebo, A.D. Patton, N.S. Rau, R.J. Ringlee, A.W. Schneider, C. Singh, L. Wang, M. Warren, P.K. Wong

May 1995

Executive Summary

This report recommends methods for pooling outage data in order to estimate indices that describe generating unit and generating system performance. The three primary recommendations of the report are summarized here.

Estimating generating **unit** performance indices requires different methods of pooling than does estimating **system** performance indices. In order to emphasize this difference, the report recommends that performance indices for system indices always include the word **system**. For example the **forced outage rate** (FOR) of a system should be called the **system forced outage rate** (SFOR).

In order to estimate unit indices, data from **homogeneous** units should be used. For homogeneous units, estimators to be pooled should be weighted inversely proportional to their variance.

In order to estimate system performance indices, the weighting of pooled estimators is chosen in order to produce an unbiased estimator.

The report describes in some detail why the recommendations are made and how to accomplish them.

Contents

Section

Tables

Appendixes

Appendix Figures

1. Introduction

Reliability and availability are important attributes of generating unit and generating system performance. Monitoring and prediction of reliability and availability are important functions in the operation, maintenance, and planning functions of electric utility organizations. This report responds to questions raised by the staff of the North American Electric Reliability Council Generating Availability Data System, NERC-GADS. The questions were related to technical and statistical procedures to assess the reliability and availability of groups of generating units. NERC maintains the Generating Availability Data System, GADS, on behalf of all US utilities and participating Canadian NERC members. Data on the performance of over 90% of the installed generating capacity in North America is maintained in GADS. These data are an important source for monitoring generating unit reliability and availability by utility organizations. A Task Force on Generating Unit Data Pooling was formed to respond to the questions. The questions were related to technical and statistical procedures to assess the reliability and availability of groups of generating units. A brief statement of the objectives set forth by the Task Force follows. This report presents the findings of the Task Force.

1.1 Task Force Objectives

The general objective of the Task Force was to identify techniques to improve estimators of reliability and availability, performance, indices of generating unit(s). The specific tasks included:

1. Identify criteria (e.g. homogeneity) for **grouping** generating units for estimating group performance trends.

2. Identify techniques (e.g. weighting, censoring of outliers) of **pooling** performance data of units in a group (e.g. a system) for estimating group performance indices.

3. Illustrate the techniques applying to estimators of different types of performance indices of generating unit(s) for different applications. There are two broad applications: monitoring performance and predicting performance.

1.2 Background

Through industry associations, electric utilities have been collecting generating unit performance data for more than 35 years. [1,2] The primary purpose of these data bases is to facilitate estimation of specific performance indices for individual units and for groups of units of common type. This effort has several applications to utility operations, maintenance and planning:

Monitoring Performance

- to compare performance of a specific unit with the average performance of similar units within the industry to determine the potential for improvement.

- to monitor performance trends and changes created by unit and/or system design, maintenance, and operational changes.

- to provide a reference for contractual capacity obligations and payments (actual performance vs a reference).

- to provide a basis for rate incentives and penalties for performance of a unit or group (system) of units (vs regional or industry averages).

Predicting Unit Performance

- to predict unit performance as input to planning studies.

- to predict performance of new units to be installed using component performance data from similar units (as input to design studies or planning studies).

- to provide a basis for defining reliability/availability standards for units.

1.3 Performance Definitions

Performance data collection is based on a standard set of classifications of generating unit operating and outage states (e.g. unplanned outage, planned outage, etc.) and deratings (e.g. unplanned, seasonal, etc.). **ANSI/IEEE Standard 762, 1987** provides the definitions of indices as well as the terms related to operations, production, outage and derated states. [3]

1.4 Data Collection Systems

World-wide, generating unit performance data is gathered by utilities, pools, regions, nations and international organizations for analyzing units as a group or groups and reporting their performance trends. NERC-GADS has been cited previously, ERIS (Equipment Reliability Information System)-Canada, UNIPEDE (Union Internationale Des Producteurs et Distributeurs D'Energie Electrique) in Europe, and ORAP (Operational Reliability Analysis Program) are examples of national and international data collection systems.

1.5 Existing Methods for Grouping Units

There are methods in current use for **grouping** units (e.g. by size range, fuel type, etc.) and for **pooling** their data to estimate the group performance (e.g. straight averages, weighting certain performance indices by unit size, etc.). These estimates of group indices are published by the data collection agencies, typically every year.

This report emphasizes procedures for grouping units and pooling the data from these grouped units. Hence, definitions of the terms **grouping** and **pooling** follow:

> **Grouping** is the process of identifying, in a data base, a set of generating units which meet specified criteria. The usual purpose of grouping units is to assemble data on generating units which are homogeneous considering one or more attributes or characteristics such as size, vintage, design, etc. for improved (i.e. smaller variance) estimation of a common parameter. In some applications, the common attribute may be common ownership or common operating control and the units are grouped for the purpose of estimating a system performance index even if the units are not otherwise homogeneous.
>
> **Pooling** is the process of aggregating the data sets from units having a known common property to improve estimates of a common parameter.

1.6 General Approach to Improving Estimators

Good estimates of generating unit performance indices are essential for operations and design evaluations and for system planning predictions. It borders on the trivial to observe that good procedures for data collection are necessary, but not sufficient; good procedures for analysis of the data are needed as well.

The general approach to improving estimators is to group units and pool their performance data. More data reduce random errors and smooth out effects of extremes (good and bad performance). Also, more data provide a large number of possibilities for classification of units, e.g. by geography, by vintage, by fuel, by manufacturer. However, bias may be introduced due to pooling data from dissimilar units, i.e. different by design or in operating and maintenance conditions. Also, estimates may vary due to different ways of pooling.

The Task Force defines improvement in estimators to mean reduction of both uncertainty and bias, and offers the observation that several techniques are available to improve estimators including: (1) increase the base of data to reduce the uncertainty by combining data from different units, i.e. pooling, (2) group units with common, i.e. homogeneous, characteristics to reduce uncertainty, (3) weight data within groups to reduce bias and also uncertainty, and (4) censor highly extraordinary data, i.e. outliers.

This report presents theoretical bases and practical approaches to improve procedures for estimating performance indices of both generating units and systems of generating units from operating and outage data. The Task Force submits that indices have been and must continue to be defined by the applications. It should also be clear that the procedures for improving estimates of performance indices must be selected specifically for the applications; there is no universal **best** procedure for all performance indices.

1.7 Performance Indices

It is useful to review the purposes and uses of specific generating unit performance indices before discussing a suitable method for pooling the generating unit performance data. For example, the Forced Outage Rate (FOR) is used widely in generation system reliability and probabilistic production cost studies. Indices including FOR, Availability Factor, (AF), and Unavailability Factor (UF), are time-based indices and depend strictly on the cumulative time in specific states. For example, FOR is based upon Forced Outage Hours and Service Hours while AF is derived from Available Hours and Period Hours. On the other hand, the Gross Output Factor, GOF, is a measure of the unit's gross generation, an energy-based index.

The criteria for grouping types of units, for pooling unit data and for weighting need not, and perhaps should not, be the same for all types of performance indices. Following this line of reasoning, the Task Force submits that performance indices can be classified into the following categories:

1.7.1 Unit Reliability and Availability Indices

These indices provide estimates of reliability and availability of a given class of generating units for productive operation. Reliability encompasses measures of the ability of a generating unit to perform its intended function. Mean Service Time to Forced Outage and Outage Frequency are typical indices of reliability (outage duration is not considered). Availability measures are concerned with the fraction of time a unit is capable of providing service, and account for outage frequency and duration. Availability Factor, Planned Outage Factor, Maintenance Outage Factor and Forced Outage Rate are used to measure availability.

1.7.2 Unit Productivity Indices

Productivity indices are concerned with the total energy produced by a unit with respect to its potential energy production; productivity indices consider magnitude of outage as well as frequency and duration of outage. The Gross and Net Output Factors belong to the category of indices which are based on actual generation. Equivalent Availability Factor also belongs to this category and is a measure of the generation that could be generated if limited only by outages and deratings.

1.7.3 System Performance Indices

System performance indices address groups of units. These indices provide general performance measures, which can be used as bases or references in maintenance planning, reliability improvement programs, etc. A plant Capacity Factor based on plant actual production and a System Equivalent Availability Factor based on unit Equivalent Availability Factors weighted by unit maximum dependable capacity are examples of performance indices developed for the specific purpose of monitoring systems of units.

1.8 Monitoring versus Prediction

While application is outside the scope of this report, recognition must be given to the role that application imposes on the choice of estimators for performance indices. There are two types of applications which should be distinguished and which can affect the choice of estimator: estimators for indices of past performance, i.e. monitors, and estimators for indices of future performance, i.e. predictors. The former is in the realm of analysis, the latter in the realm of synthesis. Both are rooted in the engineering application, but the requirements for the development of the estimators for the same indices can be quite distinct between monitoring and prediction. For example, much engineering discussion has been devoted to the benefits of different weighting methods for combining the data from a group of units including: **straight average** and **capacity weighted**. Usually, the examples and the objects are to form performance indices for dissimilar units as in a generating plant or system and the arguments turn on the effectiveness of the procedure to provide a satisfactory estimator for monitoring purposes. The discussion is appropriate to this point. Problems arise where the discussion turns to the use of the procedure for improving the estimators of groups of similar units as for example the fossil-steam size, type, age and fuel groups in the NERC GADS. It does not follow that a procedure for grouping and pooling data from homogeneous units designed to minimize bias and uncertainty of performance statistics for such a grouping bears any relationship to procedures that provide the most satisfactory estimators to be used as performance monitors for disparate units in a plant or system. For example, a plant monitor derived from the capacity weighted combination of performance data from several units should not be used to predict the effects of new unit installations with different mixes of units. Prediction requires the ability to synthesize the estimates of performance indices for the individual units. Synthesis requires the best estimators (least bias and uncertainty) for the individual units, i.e., from the constituent parts.

The Task Force submits that the proper question to ask is rooted in the application. In essence, data collection, analysis and estimation must have a purpose and defined parameter(s) to estimate. With respect to the procedure for estimating unit performance indices from groups of similar units, the procedure that minimizes both bias and uncertainty is the choice.

1.9 Report Summary

This report presents material on the following topics related to statistical methods and techniques for improvement of estimators:

1.9 Report Summary (continued)

Possible ways of pooling data including the engineering or *a priori* considerations and the pitfalls associated with improper pooling.

Methods of pooling and the associated statistical concepts, measures of goodness, homogeneity, outliers.

Sample calculations to illustrate the effects of statistical methods on sample data sets.

Conclusions.

2. Possible Ways of Pooling Data

2.1 Why Pool? Why Not?

As stated earlier, generating unit performance indices are estimated from recorded operating and outage data for two different kinds of applications: (1) monitoring performance, and (2) predicting performance. It is well known in statistics theory that the variance of an estimator is smaller for a larger sample size. A smaller variance means that the estimator has a narrower distribution and, therefore, the confidence level is higher that the estimate is closer to the **true but unknown value** of the parameter. One way of increasing the sample size is to group units and pool data from several groups of generating units. However, pooling may give misleading results if it is carried out without due consideration to such factors as unit grouping, homogeneity, and applications.

In this chapter we first consider the case where predicting unit performance is the primary application and the goal is to group units that are homogeneous in order that performance indices can be estimated for certain types of units, e.g. an estimate of the forced outage rate of 60 MW gas turbines. The chapter concludes by considering the case of pooling for monitoring performance of systems consisting of units that are evidently not homogeneous.

2.2 Pooling Data for Predicting Unit Performance

The purpose of this section is to consider the case of estimating unit performance parameters. In this case the primary goal is to estimate the parameters from homogeneous data. The inherent problem of pooling data that are not homogeneous is the interpretation of the parameters estimated from the pooled data. For example, if the availability factor were computed from the pooled operating and outage data of coal-fired baseload units and peaking combustion turbine units, it would be difficult or impossible to decide how to use the resulting estimated availability factor in a planning study. Consequently this index is not used for planning. While this is an obvious example, there may exist differences between groups of generating units that are not so obvious. Hence, test procedures are needed to ensure homogeneity in the pooling process.

2.2.1 Homogeneity

Statistically speaking, k populations are homogeneous if the distribution functions are identical [4]. Conceptually the homogeneity test of generating unit performance data can be carried out in two steps. First, the data samples to be tested must be obtained from a **nominally homogeneous** group of generating units. In order to define what constitutes a nominally homogeneous group, a set of criteria is needed which forms the basis for unit selection. This set of criteria may include both design parameters (such as fuel type, boiler type, manufacturer, etc.) and operating considerations (such as dispatch type, actual fuel burned, maintenance program, etc.). A discussion of the relevant design and engineering considerations for formulating selection criteria is given in the next section.

Operating and outage data from nominally homogeneous groups of generating units may not be homogeneous in statistical terms. This problem is resolved in the second step in which statistical tests are carried out to determine homogeneity. One possible statistical test for homogeneity is described briefly in Appendix A. Details of this method are given in [5].

2.2.2 Generating Unit Grouping Criteria

Generating units have large numbers of design and engineering parameters that may be included in grouping considerations. Typically, publications released by utility-sponsored data gathering organizations in North America have based their reports on relatively few parameters which include unit type, unit size, and primary fuel. This practice is continuing despite some concerns that other parameters may be equally important and should also be considered. A recent literature survey reveals that nearly 60 criteria had been considered in defining groups of fossil units. These criteria were ranked by their number of appearances and significance given in these publications. The results of this effort are given in Table 2.1 where the selection criteria are ranked in five levels. Obviously, there are significant differences between parameters listed in this table and the traditional parameters of unit type, size, and primary fuel.

While engineering judgment can define candidate criteria for grouping, statistical methods provide a more objective test of the validity of groups. In a recent study conducted by NERC and Southern Company Services, Inc. (SCS), a method has been developed for determining appropriate peer unit groups and for benchmarking of unit performance indices [6]. This method is briefly described in Appendix B.

It should also be noted that some of the criteria used in determining unit groupings, such as boiler type or fuel type, are of the class type and can be easily applied; others, such as unit size or vintage, are of the continuous type and must be defined by specified ranges. The ranges are usually determined by engineering judgment, or they may be statistically validated as done in the NERC and SCS procedure.

The above discussions clearly indicate that in practical applications it may not be possible to include all the design and operating considerations in the grouping criteria. In other words, the analyst may have to ignore certain (minor) aspects that are not homogeneous in a selected group of generating units. The decision may depend on the particular application and/or the performance indices to be estimated.

Table 2.1

A Sample Set of Criteria For Grouping Fossil Units (Compiled by NERC/SCS)

High Priority (#1)

- Capital and O&M Costs
- Vintage (Commercial Date)
- Loading Characteristics (Actual vs. Design)
- Quality of Preventive Maintenance Program
- Service Factor or Service Hours
- Boiler Type (Sub vs. Super Critical)
- Primary Fuel (No Lignite)

Medium/High Priority (#2)

- Boiler Manufacturer
- Turbine Manufacturer
- Oil/Coal Coal/Oil Conversion
- Scrubber Vintage
- Planned Outage Factor

Medium Priority (#3)

- Unit Nameplate (MW)
- Furnace Draft
- Balanced Draft Conversion
- Off-Spec Fuel Quality (Coal)
- Scheduled Maintenance Cycle
- Multi-Turbine/Multi-Boiler
- Furnace Bottom Type
- Scrubber & Precipitator Type
- Output & Capacity Factors
- Steam Pressure & Temperature
- Furnace Volumetric/Surface Release Rates
- Service Hours per Startup
- Use of Spares in General (Mills, BFP, ID and FD Fans, Scrubber Modules, etc.)

Low/Medium Priority (#4)

- Turbine Combustion
- Age
- Air Preheater Type
- Ash Removal System
- Inside/Outside Boiler-Turbine Installation
- Technical Add-Ons (Scrubber)
- Design or Retrofit Scrubber
- Generator Manufacturer
- BFP Drive Type
- Condenser Water Type
- Architect-Engineer
- Startup Attempt/Success Ratio

Low Priority (#5)

- Reheat (Single)
- Fuel Firing System
- Gas Recirculation Fan
- Condenser Tube Material
- Cooling Tower Type
- ID Fan Type
- Open/Closed Cooling System
- Condensing/Non-condensing Turbines
- NERC Regions

Generating unit operating and outage data usually consist of unit size (MW), period hours, the number of startups, service (or operating) hours, derated operating hours, reserve shutdown hours, forced outage hours, energy generated (MWH), and maintenance and overhaul hours, etc. In carrying out homogeneity tests, perhaps it is neither practical nor necessary to test every set. Since the majority of unit performance indices are derived from operating and forced outage hours, it might be sufficient to test these two sets only. The effect of derated operations can be represented by equivalent service and forced outage hours and included in the tests. If the operating and forced outage hours from a group of generating units test homogeneous, these units can be assumed to belong to the same population and, consequently, data pooling is permitted.

It is not uncommon that generating unit operating and outage data and, in some cases, unit performance indices contain extremely large or small values. Extreme data points in the sample, known as outliers, may unduly influence (bias) the analysis. Should they be included in or excluded from the analyses? What factors should be considered? What are the options? These questions are discussed in some detail in Section 3.3.

2.3 Pooling Data for Monitoring System Performance

Reliability and productivity indices have been used for many years to compare individual unit performance against a group yardstick and for performance trending as diagnostics to indicate need for maintenance and overhaul. A recent practice has been to develop performance measures for systems, this requires the formation of reliability and productivity indices for collections of heterogeneous units. Most recently, indices have been used as a part of Performance Evaluations for independent suppliers of energy to establish rate incentives and for regulators to establish incentives for utilities. An example may be drawn from recent actions of a Public Service Commission [7]. Performance evaluations for the purpose of establishing rates and return on equity were based on a weighting of several attributes including Customer Service Reliability and a System Equivalent Availability Factor, SEAF. Benchmarks for each may be established in terms of utility, area and national performance. SEAF is a performance index derived from the Equivalent Availability Factors, EAF, for a heterogeneous mix of fossil-steam units of mixed sizes, fuels and boiler types. SEAF is defined as the average of the EAF's for fossil-steam electric generating units for the prior twelve month period.

SEAF	$= \sum_{j=1}^{N} [EAF_j]^{*}[1/N]$
EAF_j	= (AH - (EPDH+EUDH+ESEDH))/PH
EAF	= Equivalent Availability Factor for period
SEAF	= System Equivalent Availability Factor for period
N	= No. of fossil units on system during period
PH	= Period hours
EPDH	= Equivalent Planned Derated Hours for unit j
EUDH	= Equivalent Unplanned Derated Hours for unit j
ESEDH	= Equivalent Seasonal Derated Hours

The performance indicator SEAF measures the average percentage of time that the fossil steam electric generating units of the system were ready and available to produce electricity during the twelve month period ending with the evaluation date. This opens a question about interpretation of the System Equivalent Availability Factor as a measure of the system capability to supply load given that unit capability is not factored into the computation of the index. Strictly, if it were desired for the index to reflect the system capability to supply load, then the equivalent availability of each generating unit should be weighted by its maximum dependable capability:

$$WSEAF = \{\sum_{j=1}^{N}[EAF_j]^{*}[MDC_j]\} / \sum_{j=1}^{N}[MDC_j]$$

WSEAF = Weighted System Equivalent Availability Factor

MDC_j = Maximum Dependable Capability for Unit j

It would appear that a capacity weighting to reflect system capability to serve load would be a preferable system index to the unweighted, arithmetic average.

Let us compare the measures SEAF and WSEAF for the Favorite Light and Power Company's fossil-steam generating units. FLAPCO's availability data for 1991 are offered in the following table.

UNIT NO	1	2	3	4	5
FUEL	Oil	Coal	Coal	Oil	Oil
MDC MW	400	800	400	300	300
AH HRS	7000	6000	6500	7500	7600
EUDH HRS	400	500	800	500	300
EPDH HRS	0	500	0	0	0
PH HRS	8760	8760	8760	8760	8760
EAF %	75.3	57.1	65.1	79.9	83.3

FLAPCO

SEAF = 72.14%

WSEAF = 68.54%

Comparing, SEAF measures the equivalent availability of the average FLAPCO unit while WSEAF measures the equivalent availability of FLAPCO generation. SEAF is the percent time the average unit is available to generate while WSEAF is the equivalent percent time that the system can generate energy. For the FLAPCO example, SEAF would slightly overstate the system capability to generate.

2.3.1 Statistical Considerations

The primary statistical consideration concerning generating unit data pooling for parameter estimation when the goal is system performance comparison becomes finding an unbiased estimator when the population is not homogeneous.

The difference between the pooled estimator from a homogeneous population and nonhomogeneous one arises from the fact that in the former case, the individual subpopulations share the common value of the parameter being estimated while in the latter case, a new composite parameter is being defined by the analyst as a function of the individual subpopulation parameters.

For example, the availability factor of a nonhomogeneous group of units (e.g. a system) may be defined as (note that the S is added as a prefix to AF to indicate that this is a system parameter)

$$SAF = \frac{N_1}{N} AF_1 + \frac{N_2}{N} AF_2 + \cdots + \frac{N_k}{N} AF_k$$

where AF_i is the parameter of subpopulation i, and N_i and N are, respectively, the sizes of subpopulation i and the total population.

An **unbiased estimator** of SAF is

$$S\hat{A}F = \frac{N_1}{N} \hat{AF}_1 + \frac{N_2}{N} \hat{AF}_2 + \cdots + \frac{N_k}{N} \hat{AF}_k$$

where $\hat{AF}_i$ is an unbiased estimator of AF_i.

In the above example the parameters of the subpopulations are weighted by N_i/N. When defining unit productivity indices, for example, the analyst may decide to use unit size as a weighting factor. After the productivity index is defined, the purpose of the weighting is simply to find an unbiased estimator of the defined index.

3. Statistical Concepts Associated with Data Pooling

3.1 Introduction

As pointed out in earlier sections, application of indices has an important role in the choice of factors that affect pooling of data. Applications, for this purpose, may be broadly classified as either planning or monitoring of performance. Homogeneity of data is important when the indices are to be used for predictive purposes as is the case in planning studies. When the indices are used for monitoring or comparison of groups (e.g. system) performance, they may be obtained from nonhomogeneous samples and apply to nonhomogeneous populations. This chapter considers the methods for pooling data and associated statistical concepts for both homogeneous and nonhomogeneous populations.

The objective of pooling data can be stated as follows. Assuming, the true, but unknown parameter or index is designated as Θ, the objective is to combine n estimators, $\hat{\Theta}_1, \hat{\Theta}_2, \cdots, \hat{\Theta}_n$ into a **best estimator** of Θ. As an example consider n estimators of the availability factor, AF, of a certain population of generators and call them $\hat{AF}_1, \hat{AF}_2, \cdots, \hat{AF}_n$. The objective then is to combine $\hat{AF}_1$ through $\hat{AF}_n$ to arrive at the best estimate of AF for the specified population of generators. Another important consideration in pooling data is the recognition and treatment of outliers. Difficulties caused by outliers and tests for their identification are discussed.

3.2 General Statistical Concepts for Pooling Estimators

3.2.1 General Form of Pooled Estimators

In any discussion of data pooling, care must be taken to see whether the population is homogeneous or nonhomogeneous. In the case of a homogeneous population, the individual subpopulations share the common value of the parameter being estimated. As an example of homogeneous populations, the individual estimates $\hat{AF}_i$, of AF, may have different variances but the same expectation. In the nonhomogeneous case the analyst defines a new composite parameter as a function of the individual subpopulation parameters. This case frequently occurs when a system parameter is to be estimated. The definition of this parameter may not involve any ideas from the theory of statistical estimation but could be based on rational judgement. As an example, the availability factor for a system, i.e. a nonhomogeneous population, can be defined as,

$$AF = \frac{N_1}{N} AF_1 + \frac{N_2}{N} AF_2 + \cdots + \frac{N_k}{N} AF_k \tag{1}$$

where AF_i is the parameter for generator type i, or subpopulation i, and N_i is the number of generating units in the i th subpopulation, $\sum N_i$-N, the number of generating units in the system.

The composite parameter can be defined in many ways but this chapter will be limited to discussion of pooling considerations to a weighted average. The pooled estimator for Θ will, therefore be restricted to the linear form given by Equation (1).

In the case where each estimator is equally weighted this reduces to the simple average,

$$\hat{\Theta}-\frac{\sum \hat{\Theta}_i}{n} \qquad (2)$$

That is, each weight is simply $\frac{1}{n}$.

The next section describes the criteria for choosing the weights in Equation (2) such that the pooled estimator is the **best estimator** of the index assuming that the population is specified.

3.2.2 Criteria for Good Estimators

Generally, by a good estimator we mean one with its distribution concentrated near the population parameter being estimated. The well accepted criteria are (1) unbiasedness, (2) consistency, and (3) efficiency.

Unbiasedness:

An estimator, $\hat{\Theta}$, is an unbiased estimator of parameter, Θ, if and only if

$$E[\hat{\Theta}]-\Theta$$

where E indicates expected value.

Since $\hat{\Theta}$ will vary from one sample to another, $\hat{\Theta}$ cannot generally equal Θ, thus it seems reasonable to choose $\hat{\Theta}$ such that it is equal to Θ on the average. In other words if the expected value of $\hat{\Theta}$ is equal to Θ, then $\hat{\Theta}$ is said to be an unbiased estimator, and otherwise it is said to be biased.

Consistency:

When the sample size is increased one would expect the estimator to come closer to the true value. An estimator is said to be consistent if the estimator approaches (with probability one) the population parameter being estimated as the sample size increases.

Efficiency:

If two estimators, $\hat{\Theta}_1$ and $\hat{\Theta}_2$, are both unbiased estimators of Θ, then it seems reasonable that one chooses the estimator that deviated from Θ the least in some expected or average sense. It is natural to choose to measure the expected deviation from the mean by the variance of the estimator. Thus $\hat{\Theta}_1$ is a better estimator than $\hat{\Theta}_2$ if the variance of $\hat{\Theta}_1$ is less than the variance of $\hat{\Theta}_2$. In general the estimator with smallest variance is said to be an **efficient estimator**.

It may be concluded from the above discussions that the **best linear estimator** is an unbiased minimum variance estimator.

3.2.3 Examples of Unbiased Minimum Variance Estimators

Estimator for the Mean

Let $X_1, X_2, \cdots, X_N$ be a population of size N. Then by definition the population mean is,

$$\mu = \frac{1}{N} \sum_{i=1}^{N} X_i \tag{3}$$

Let $X_1, X_2, \cdots, X_n$ be a sample of size n, where $n < N$. The sample mean is

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{4}$$

It can be easily shown that

$$E(\overline{X}) = \mu$$

That is, the sample mean is an unbiased estimator of the population mean.

The variance of $\overline{X}$

$$V(\overline{X}) = \frac{\sigma^2}{n} \tag{5}$$

where $\sigma^2 =$ population variance. We know from the Cramer-Rao inequality [8] that $\frac{\sigma^2}{n}$ is the smallest variance of an estimator of $\overline{X}$. Therefore the sample mean is an unbiased and least variance estimator of the population mean.

Estimator of Availability Factor

Availability factor, AF, may be considered to be a special case of the mean $\overline{X}$ where the variable, X_i takes on either 0 or 1 value. For example consider,

$$AF = \frac{AH}{PH} \tag{6}$$

Let there be N number of period hours, i.e.,

$$N = PH$$

and let $X_i = 1$ when the generator is available and 0 when it is unavailable, then

$$AH = \sum_{i=1}^{N} X_i \tag{7}$$

and

$$AF = \frac{AH}{PH} = \frac{\sum_{i=1}^{N} X_i}{N} \tag{8}$$

It is to be noted that Equation (8) for availability is the same as the Equation (3) for mean except that variable X_i takes on values 0 or 1 only. Using the considerations for the estimator of $\overline{X}$, it can be shown that the availability estimator,

$$\hat{AF} = \sum_{i=1}^{n} \frac{X_i}{n} \tag{9}$$

when X_i = 0 or 1, is an unbiased estimator of AF.

A derivation of the mean and variance of $\hat{AF}$ for a two state generating unit which is operating continuously, assuming constant failure and repair rates, is given in Appendix C.

3.2.4 Best Pooled Estimators Using Samples from a Homogeneous Population

Assume that there are n independent estimators $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_n$ of θ and that each estimator is unbiased. It is also assumed that the ith estimator, $\hat{\theta}_i$, has a variance σ_i^2, $i = 1, \cdots, n$. A pooled estimator of the following form is required

$$\hat{\theta} = \sum_{i=1}^{n} \omega_i \hat{\theta}_i \tag{10}$$

such that $\hat{\Theta}$ is unbiased and minimum variance.

It is easy to show that the criterion of unbiasedness requires

$$\sum_{i=1}^{n} \omega_i = 1 \tag{11}$$

and the criterion of minimum variance requires, given that the $\hat{\Theta}_i$'s are independent,

$$\omega_i = \frac{\frac{1}{\sigma_i^2}}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}} \tag{12}$$

Equation (12) used in Equation (10) defines the best pooled estimator. The examples below illustrate the application of these criteria in the case of five parameters from homogenous populations.

In all five examples the population is assumed to be homogeneous; the sample is assumed to be representative of the population; and the different generating units are assumed to be independent. The data are from Table 3.1. With these data the objective is to estimate five indices of coal-fired steam units for which units numbered 1 through 3 are assumed to be a representative sample. We want to pool the estimators derived from these three units. In other words, we wish to find ω_1, ω_2, and ω_3 in the equation.

$$\hat{\Theta} = \omega_1 \hat{\Theta}_1 + \omega_2 \hat{\Theta}_2 + \omega_3 \hat{\Theta}_3$$

such that $\hat{\Theta}$ is unbiased and minimum variance. The five examples differ only in the parameter being estimated.

Example 1 - Estimate AF (Availability Factor)

It is shown in Appendix C that $\hat{AF}_i$ is unbiased and that the variance of $\hat{AF}_i$ is approximately proportional to the reciprocal of period hours. Thus using Equation (12)

$$\omega_i = \frac{PH_i}{PH_1 + PH_2 + PH_3}, \ i = 1,2,3.$$

The estimate of AF is thus

$$\hat{AF} = \frac{8760(.80)+8760(.68)+6000(.67)}{(2)8760+6000} = \frac{7000+6000+4000}{(2)8760+6000} = \frac{\sum \text{Available Hours}}{\sum \text{Period Hours}} \approx .722 \text{ or } 72.2\%$$

Example 2 - Estimate FOR (Forced Outage Rate)

The individual estimators, $\hat{FOR}_i$, are unbiased and the variance is approximately proportional to reciprocal of observation time, i.e.

$$Var[\hat{FOR}_i] \approx \frac{k_1}{SH_i+FOH_i},\ i=1,2,3$$

where k_1 is a constant.

Thus

$$\hat{FOR} = \frac{7500(6.67)+7000(14.29)+4500(22.22)}{7500+7000+4500}(100) = \frac{500+1000+1000}{7500+7000+4500}(100) = \frac{\sum \text{Outage Hours}}{\sum (\text{Outage Hours}+\text{Service Hours})} \approx 13.2\%$$

Example 3 - Estimate EFOR (Effective Forced Outage Rate)

The variance of the estimator of effective forced outage rate is approximately inversely proportional to observation time. Thus

$$Var[E\hat{FOR}_i] \approx \frac{k_2}{SH_i+FOH_i+ERSFDH_i}$$

where k_2 is a constant.

Thus

$$\hat{EFOR} = \frac{7800(11.54)+7000(14.29)+4900(36.73)}{7800+7000+4900}(100)$$

$$= \frac{500+400+1000+1000+800}{7800+7000+4900}(100)$$

$$= \frac{\sum (FOH+EFOH+ERSFDH)}{\sum (\text{Outage Hours}+\text{Service Hours}+\text{Equivalent Reserve Shutdown Forced Derated Hours})} = 18.8\%$$

Example 4 - Estimate CF (Capacity Factor)

In Appendix D it is argued, using some broad assumptions, that $\hat{CF}_i$ is unbiased for CF and that $\hat{CF}_i$ has a variance proportional to the period hours, PH_i.
Thus

$$\hat{CF} = \frac{8760(.63)+8760(.68)+6000(.54)}{(8760)(2)+6000}(100)$$

$$= \frac{2200+2400+1300}{400[8760(2)+6000]}(100)$$

$$= \frac{\sum \text{Actual Generation}}{\text{Capacity} \sum \text{PH}} = 62.7\%$$

Example 5 - Estimate MSTFO (Mean Service Time to Forced Outage)

If there are r observed failures, then $\frac{2r\hat{MSTFO}}{MSTFO}$ will be approximately chi-squared distributed with $2r$ degrees of freedom. Thus

$$Var[\hat{MSTFO}_i] = \left(\frac{MSTFO}{2r}\right)^2 Var[\chi^2_{2r}]$$

$$= \frac{(MSTFO)^2}{4r^2}(4r) = \frac{(MSTFO)^2}{r}$$

Thus the variance of $\hat{MSTFO}_i$ is approximately inversely proportional to the number of failures, and we have

$$\hat{MSTFO} = \frac{(3)2333.33+(4)1500+(6)583.33}{13}$$

$$= \frac{7000+6000+3500}{13} = 1269$$

$$= \frac{\sum SH}{\sum \text{Number of Unplanned Outages}}$$

Table 3.1

Example Generator Data for Estimating Pooled Parameters

Unit Number	1	2	3	11	12	21
Unit Type	[Coal	Fired	Steam]	[Gas	Fired Steam]	C.T.
Unit Size (MW)	400	400	400	300	300	50
Period Hours (PH)	8760	8760	6000	8760	7000	8760
Available Hours (AH)	7000	6000	4000	7500	5000	5000
Service Hours (SH)	7000	6000	3500	7000	5000	500
Forced Outage Hours (FOH)	500	1000	1000	600	400	400
Equivalent F.O. Hours (EFOH)	400	0	800	500	300	0
ERSFDH	300	0	400	200	300	0
Number Unplanned Outages	3	4	6	2	3	5
Actual Generation (AAG) (GWH)	2200	2400	1300	1900	1500	22

Indices for the Individual Units

Availability Factor (AF) - %	79.9	68.5	66.7	85.6	71.4	57.1
Forced Outage Rate (FOR) - %	6.7	14.3	22.2	7.9	7.4	44.4
Equivalent FOR (EFOR) - %	11.5	14.3	36.7	14.1	12.3	44.4
Capacity Factor (CF) - %	62.8	68.5	54.2	72.3	71.4	5.0
Mean Service Time To Forced Outage (MSTFO)	2333.33	1500.00	583.33	3500.00	1666.67	100.00

3.2.5 Best Pooled Estimators for a Nonhomogeneous Population

If there are k generating unit types that make up the population (system) for which the index is to be estimated, then the population is called nonhomogeneous. For example, the availability factor, *AF*, of the entire generating system of Table 3.1 might be of interest. The generating system consists of three 400 MW coal fired generating units, two 300 MW gas fired steam units, and one 50 MW combustion turbine, and we would call this an nonhomogeneous population.

The availability factor estimated from a nonhomogeneous sample may be of little value for use in planning studies, but it can be useful for monitoring of performance. If this is the quantity to be estimated then the criteria of unbiasedness is the primary criteria. If in the population there are N_1 generating units from the first homogeneous subpopulation, N_2 generating units from the second homogeneous subpopulation, etc., and finally N_k units from the k th homogeneous subpopulation, then the total nonhomogeneous population consists of

$$N=N_1+N_2+\cdots+N_k \text{ generating units}$$

If $\hat{AF}_i$ is the estimator of the availability factor of the ith subpopulation, $i=1,\cdots,k$, then it can be shown that the unbiased estimator of SAF (system availability factor) is

$$\begin{aligned} S\hat{A}F &= \frac{N_1}{N}\hat{AF}_1+\frac{N_2}{N}\hat{AF}_2+\cdots+\frac{N_k}{N}\hat{AF}_k \\ &= \sum_{i=1}^{k}\frac{N_i}{N}\hat{AF}_i \end{aligned} \qquad (13)$$

Continuing the example described above using the data from Table 3.1 and using Equation (13) to estimate the system availability factor, $S\hat{A}F$.

$$\begin{aligned} S\hat{A}F &= \left\{\frac{3}{6}[.722]+\frac{2}{6}\left[\frac{7500+5000}{8760+7000}\right]+\frac{1}{6}[.57]\right\}(100) \\ &= 72.1\% \end{aligned}$$

where ".722" is calculated in Example 1.

3.3 Outliers in Generating Unit Data Pooling

3.3.1 What to Do About Outliers

An outlier in a set of data is a sample (or a subset of samples) which **appears to be inconsistent** with the remainder of the sample population. It can be either very large or very small in comparison with the main body of samples. The presence of outliers raises the following questions: (1) Are the outliers genuine members of the main population, or alternatively, are there errors in measuring, recording or transmitting the data? (2) Do they cause difficulties in fitting the population to a probability model? (3) In the case that no probability model representation of the data set is sought, are the outliers significantly distorting the estimates of parameters to make us believe that they do not belong to the main population?

Outliers may appear in the recorded data on the operating or outage times of a generating unit or in the computed indices for generating units:

1. One or more entries of outage times may appear to be extremely long in comparison with the rest of the outage data for a given generating unit. The long outage times are commonly caused by unusual failure events of major equipment. In many cases, a serious equipment failure traced to inadequate design may also lead to long planned outages of other units of identical design to replace the equipment in question so as to avoid the occurrences of the same type of failure. On the other hand, recorded entries of long operating times may either indicate exceptional performance or non-compliance with reporting procedures.

2. One or more of the performance indices of a given generating unit (or units) is very low (high) in comparison with those for the rest of the units in the data pool. The causes of poor performance indices can be either unusual operating circumstances or serious equipment failures during the period of interest, while high performance indices could be due to usual operating circumstances or poor reporting procedures.

Note that the presence of outliers in these two places may not be correlated. In other words, the presence of an outlier in the outage times of a given generating unit may not cause the computed performance indices of this unit to become outliers among indices for all the units in the data pool.

The problem of how to carry out the data pooling process in the presence of outliers will be examined pertaining to the questions outlined at the beginning of this discussion. Also, the discussion will be limited to outliers in outage data only. Outliers of operating data can be treated in a similar manner.

It has been a common practice in most outage data systems to have built-in checking mechanisms in the data entry process so that errors or inconsistencies will be exposed and corrected. Therefore, we assume that there are no errors in the outage data to be pooled. In other words, the recorded outage data are all members of the sample population. Therefore, the rejection of an outlier depends on the need for a probability model and on whether it will significantly distort parameter estimates.

In the case where a probability model of the data samples is required, serious considerations should be given to the exclusion of outliers in order to obtain the best fit. For example, in generating system reliability studies, the exponential distribution is often chosen to represent the uptimes and downtimes of generating units for reasons of simplicity and tractability in the mathematical analysis. One would have a strong argument to reject the outliers in the outage data for use in such studies. Otherwise it may be necessary to select a different distribution which would be much more difficult to handle in the mathematical analysis. Results of recent investigations have shown that, for independent generators, the choices of outage data distribution models do not have significant effect on the expected values of the reliability assessments [9].

When a probability model is not required, testing and rejection of the outliers can be considered only if they significantly distort parameter estimations. However, the sample sizes of pooled data of generating units are expected to be large, and the effect of the outliers on the average values of the computed indices is likely to be insignificant. In the case where a particular set of outliers is attributable to a cause which is known to be unlikely to recur, then these outliers can be rejected in computing the performance indices of an **average unit**.

The next question to ask is: Should the procedure of detecting outliers be carried out for individual units before data pooling, or should it be carried out for the pooled data? If the outlying data samples of given generating units are known to be caused by some rare events, they should be scrutinized before pooling. In the case where probability modeling is required, it seems to be logical to pool the data first before testing for outliers. Obviously these are conflicting suggestions. However, the major concerns of the Task Force on Data Pooling are the selection of minimum variance, unbiased estimators for computing weighted average performance indices of generating units and the formation of industry-wide standards. From these points of view, it is easy to decide that outlying data samples should be examined before they are submitted to the pool.

To summarize, it is a subjective judgment on the part of the analyst whether or not to exclude the outlying data samples in the analysis process. With regard to the concerns of this task force, it is sufficient to suggest the following general guidelines.

1. Testing for outliers is recommended in cases where probability modeling of data is required.

2. Outliers may be examined and subsequently excluded from the data pooling process if they are known to be caused by rare events, or if they can cause significant distortions in parameter estimations (or in the computation of performance indices).

3. Examination of outlying data should be made for individual units before pooling.

3.3.2 A Test for Outliers

We describe a test for upper outliers. A simple method for such tests requires the assumption that the data samples under examination follow a normal distribution with unknown mean and variance. The test procedures are as follows:

Arrange the data samples in an ascending order, denoted by $x_1, x_2, \cdots, x_n$.

1. Test for a single upper outlier:

(1) Compute the test statistic $T = (x_n - \bar{x})/s$, where $\bar{x}$ is the sample mean and s is the sample standard deviation.

(2) Select an appropriate significance level (usually, either 5% or 1%).

(3) Compare the value of T with the corresponding entry in a table for outlier testing (e.g., Table VII of Reference [7]), and reject the outlier if the value of T is larger than the table entry.

2. Test for k upper outliers:

(1) Compute the test statistic $T = (x_{(n-k+1)} + \cdots + x_n - k\bar{x})/s$.

(2) Select an appropriate significance level.

(3) Compare the value of T with the corresponding entry in a table for multi-outlier testing (e.g., Table IXa of Reference [10]), and reject the k samples as outliers if the value of T is larger.

4. Sample Calculations

Consider the data from a system composed of the units described in Table 4.1. How can these data be "pooled" to estimate EAF and FOR in accordance with the recommendations of this report?

Strictly speaking, EAF and FOR are measures applied to **units** to be used for **prediction** and only data from homogeneous units should be pooled. Without a detailed investigation of homogeneity as described in Chapter 3, we assume that Units 4 and 5 can be grouped in a homogeneous group and their data can be pooled. In this case because the period hours (PH) are the same for both units, their EAF's are equally weighted (see Section 3.2.3 of the report) and for 300 MW oil fired units

$$\hat{EAF} = \frac{79.91 + 83.33}{2}(100) = 81.62\%$$

In general, data would be weighted inversely proportionally to period hours (see Section 3.2.3).

In order to estimate the FOR for these homogeneous units the weighting should be proportional to the reciprocal of observation time (Section 3.2.3 and Appendix C show that the variance of this estimator is inversely proportional to the observation time), i.e. SH + FOH; thus for 300 MW oil fired units

$$\hat{FOR} = \frac{8410(34.6) + 8660(41.11)}{8410 + 8660}(100)$$

$$= \frac{(2910 + 3560)(100)}{5500 + 2910 + 5100 + 3560} = 37.9\%$$

While it is possible that Unit 1 might be grouped into a homogeneous group with Units 4 and 5, the investigation of homogeneity is beyond the scope of these sample calculations. However the procedure to decide about homogeneity is described in Section 2.2.1 of this report.

We now consider estimating **system** parameters to be used in **performance** comparisons. In this case, this report implies:

* Unit parameters, e.g. EAF and FOR, **should not** be estimated from data pooled from non-homogeneous groupings.

* The analyst (e.g. regulator, economist, utility system monitor) can define a system parameter such as SEAF or WSEAF (see Section 2.3 of the report), and the parameter can be estimated in order to arrive at a system performance measure.

$$SEAF = \frac{1}{N} \sum_{j=1}^{N} EAF_j$$

Then for the system from Table 4.1

$$SE\hat{A}F = \frac{1}{7} (75.34+57.08+65.07+79.91+83.33+91.21+5.00)(100)$$

$$= 65.28\%$$

$$WSEAF = \sum_{j=1}^{N} \frac{(EAF_j)(MDC_j)}{\sum_{i=1}^{N} MDC_i}$$

Then for the system from Table 4.1

$$WS\hat{E}AF = \frac{1250(91.21)+800(57.08)+400[75.34+65.07]+300[79.91+83.33]+30(5.00)}{1250+800+2(400)+2(300)+30}(100)$$

$$= 76.14\%$$

Other weighted system performance measures such as NSEAF (New System Equivalent Availability Factor) could be defined. Note that it is our strong recommendation that **"S"** be included in the defined parameter in order to emphasize that it is a **system** parameter.

* System parameters should not be used to apply to a partial system, e.g. Units 6 and 7 of the sample system given in Table 4.1.

Finally, it should be emphasized that it is beyond the scope of this report to define which system parameter is to be defined to measure system performance. Furthermore we believe that the choice of parameter to measure system performance is not totally an engineering question, but other considerations such as economics and politics must be considered.

Table 4.1

Data of a Generating System for Sample Calculations in Pooling

Data

Unit Name	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	Unit 6	Unit 7
Fuel	Oil	Coal	Coal	Oil	Oil	Nuclear	C-T
MDC MW	400	800	400	300	300	1250	30
AH Hrs	7000	6000	6500	7500	7600	8000	20
EUDH Hrs	400	500	800	500	300	10	0
EPDH Hrs	0	500	0	0	0	0	0
SH Hrs	6800	6000	4000	5500	5100	8000	3
PO Hrs	744	100	744	350	100	0	150
FO Hrs	1216	2660	4016	2910	3560	760	247
PH Hrs	8760	8760	8760	8760	8760	8760	400

EAF%

Unit Level	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	Unit 6	Unit 7
EAF%	75.34%	57.08%	65.07%	79.91%	83.33%	91.21%	5.00%

FOR%

Unit Level	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	Unit 6	Unit 7
FOR%	15.17%	30.72%	50.10%	34.60%	41.11%	8.68%	98.80%

5. Conclusions

Techniques for reviewing operational histories of individual electric generating units have been well defined and readily applied in the utility industry. Efforts by IEEE and other international technical organizations have led to standard terms, definitions, and calculation procedures for unit performance indices.

Two primary applications of operational data are for:

1. Predicting or estimating parameters of unit performance to be used for planning new unit additions to the electric system, operating existing units, and designing equipment for new or existing units.

2. Estimating system performance indices to be used for supporting Regional evaluations, and management or regulatory reviews.

This report recommends methods for pooling outage data from **homogeneous** units for predicting unit parameters. Homogeneous units have similar design and operational characteristics such as vintage, primary fuel, manufacturer, and capacity ratings. Testing to determine if the pooled group of units is homogeneous is recommended in the report. This report discusses methods for predicting performance indices such as Equivalent Availability Factor and Forced Outage Rate. One general result of the report is that for homogeneous units, estimates to be pooled should be **weighted inversely proportional to their variance**.

Monitoring the performance of a group of units that constitute an electric system or a sub-set of a system is a common practice in the electric utility industry. Choosing which parameter to use to measure system performance and which units to include may not be an engineering or an analytical decision. Such a system is very likely to be comprised of units that are not homogeneous. If so, this report recommends that the estimation of indices for non-homogeneous groups be done using a weighting technique and the **weighting is chosen to produce an unbiased estimate** of the defined performance parameter.

To clearly denote that the system is non-homogeneous, and that the index may not represent an industry-wide group of units, the report recommends that indices for non-homogeneous groups include a reference to the term **system**. An example of this would be that the Forced Outage Rate would be represented as the **System** Forced Outage Rate.

This report is designed to clearly delineate the two major reasons for pooling estimators, and to recommend the correct method for pooling in both cases.

6. References

[1] Edison Electric Institute "Report on Equipment Availability for the Seven Year Period 1960-66", EEI Publication 67-23, 1967.

[2] North American Electric Reliability Council, Generating Availability Data System, Annual Reports, Princeton, N.J., U.S.A..

[3] Institute of Electrical and Electronics Engineers, IEEE Standard Definitions for Use in Reporting Electric Generating Unit Reliability, Availability and Productivity, ANSI/IEEE Std. 762, 1987.

[4] G. James and R.C. James (edit), *Mathematics Dictionary,* Van Nostrand, Princeton, N.J., 1949.

[5] A.H. Chan and F. Camacho, "K-Sample Test of Homogeneity", unpublished technical report of Ontario Hydro Research Division, presented at the Statistical Society of Canada Annual Meeting, Ottawa, 1989.

[6] "Report, Target Unit Availability Review", NEPOOL Unit Availability Task Force, June, 1992.

[7] Mississippi Public Service Commission "Performance Evaluation Plan Rate Schedule PEP-1" Issued by H.E. Blakeslee, Vice President, Mississippi Power Co., Gulfport, Miss., Dec. 28, 1990.

[8] J.E. Freund and R.E. Walpole, *Mathematical Statistics,* 4th Edition, p. 335.

[9] C. Singh, "Effect of Reliability Distributions on Steady State Frequency", IEEE Trans. on Reliability, Vol. R-29, No. 3, August 1980, p. 274.

[10] V. Barnett and T. Lewis, *Outliers in Statistical Data,* John Wiley and Sons, Chichester, England, (1978).

APPENDIX A

A Homogeneity Test Method

Tests of homogeneity are used to determine whether observations from different populations have the same distribution. For the case of two populations the Kolmogorov-Smirnov two-sample test is commonly employed. This statistic is difficult to apply to more than two populations. The SAS uses an ad-hoc procedure to solve this problem by using the sum of the squared differences between the empirical distribution function (EDF) and the pooled one (see SAS/STAT Guide Version 6, P. 718, 1987). However, this approach does not account for the dependency among the squared differences. A new k-sample test method has been developed which is easy to apply and rectifies the problem encountered in the SAS test. This method is briefly described in this Appendix. A sample application to test the homogeneity of operating and outage data of generating units is also presented.

The k-sample test of homogeneity proposed in [5] is based on comparisons between the empirical distribution function (EDF) of each of the sampled populations with an estimate of the hypothetical common distribution. The common distribution is estimated using a data set obtained by randomly selecting half of the observations from each of the populations.

Let

n_i =sample size of the i th population P_i, $i=1,2,...,k$.

x_{ij} =j th sample of P_i, $j=1,2,...,n_i$ and $i=1,2,...,k$.

n =Sum$_i(n_i)$.

Define

$$F_{n_i}(x)=\frac{Number\ of\ \{x_{ij} \le x \mid x_{ij} \in P_i\}}{n_i}$$

$$F_{\frac{n}{2}}(x)=\frac{Number\ of\ \{x_{ij} \le x \mid x_{ij} \in P_{\frac{n}{2}}\}}{\frac{n}{2}}$$

$$F_n(x)=\frac{Number\ of\ \{x_{ij} \le x \mid x_{ij} \in P_n\}}{n}$$

where $P_{n/2}$ is the data set obtained by randomly selecting half of the observations from each of the populations and P_n is the total pooled population.

The test statistic, D, is defined as

$$D = sup_x \| F_{n_i} (x) - F_{\frac{n}{2}} (x) \|$$

The application of this homogeneity test method was illustrated in an example in which the uptimes and downtimes of 4 nominally identical coal-fired units were tested. The sizes of the data samples are as follows:

Unit #	Sample Sizes of Uptime	Sample Sizes of Forced Outage Time
1	960	66
2	879	76
3	762	70
4	849	75

Note that the sample sizes of uptimes and forced outage times for each unit are not equal because of deratings and scheduled outages.

Following [5], a generalized Kolmogorov-Smirnov statistic, D, based on the EDF for half of the pooled data, were computed for 500 random selections of half of the observations of each population. The critical region of size 0.05 was selected which corresponds to 95% probability of not making an incorrect decision. Accordingly, the critical level D_0 equals 3.285 and the homogeneity hypothesis will not be rejected if D is smaller than 3.285, and rejected otherwise.

For the data on the forced outage times the computed values of D are lower than D_0 for all random half samples, with the highest value of D=2.404. For the uptimes, the computed value of D are also lower than D_0, with the highest value of D=2.935. These results indicate that the generating unit uptimes and forced outage times are homogeneous.

The estimated distribution function using the entire samples of all k populations can also be used. In this case the statistic to be used is

$$sup_x \| F_{n_i} (x) - F_n(x)\|$$

The sensitivity of this test will be less than the one using the randomly selected half of the pooled data.

APPENDIX B

An Approach to Determine Unit Groupings

This appendix describes a study, conducted by NERC and SCS [6], in which a method has been developed for determining appropriate unit groupings based on design and operational data. The purpose of determining the unit groupings is to prepare benchmarks of performance indices for steam generating units. The performance indices to be benchmarked include:

1) Reliability - Equivalent Forced Outage Rate (EFOR)
2) Availability - Equivalent Availability Factor (EAF)
3) Maintainability - Scheduled Outage Factor (SOF)

Benchmarking is very important in today's competitive environment. Utilities must be able to assess how a unit's performance compares to its peers so that aggressive, yet achievable cost-effective goals for the generating units can be set. These goals can help local plant management to recognize the need and opportunities for performance improvement.

The proposed approach utilizes GADS database which contains over 30 years of data on over 4,000 North American generating units.

Numerous design/operation factors were tested statistically to find the most appropriate peer group for each unit. Once a peer group for a given unit was found, the unit's overall performance was compared against the distribution for the entire group.

Statistical comparisons were made of the distributions of unit performance parameters resulting from design and operating factors including unit size, fuel type, boiler criticality, boiler/turbine manufacturer, vintage, etc. The parameter which showed the largest statistical difference was chosen as the primary criterion for unit grouping. This process is repeated until all the parameters have been examined.

The EFOR index was used in the analysis to determine which design and operation factors were the most important in selecting peer unit groups.

The first step was to perform a normality test on the variables EFOR and XEFOR. The variable XEFOR is simply an arc sine transformation of the variable EFOR. This transformation is performed to convert non-normal distributions into normal ones so that a more powerful parametric test (analysis of variance) can be used.

The normality test results determine whether parametric or nonparametric significance tests would be most appropriate.

If the normality test shows that the sample data can be considered normal, then parametric procedures based on a normal distribution can be used. The method of least squares is used to fit linear models.

A nonparametric procedure is used if the normality test shows that the data cannot be considered normal (after the arc sine transformation). This procedure performs an analysis of variance on the rank scores of a response variable ($EF\hat{O}R$) across a one-way classification. The rank scores are functions of the ranks of the response variable ($\hat{EFOR}$), where the values are ranked from low to high. The procedure also calculates linear rank statistics based on WILCOXON scores. These statistics are used to test if the distribution of a variable has the same location parameter across different groups.

In some instances it may be desirable to substitute EAF or SOF as the test variable.

The flowchart in Figure B.1 summarizes the analysis process.

The engineering and operating characteristics which are tested for significance in the analysis process are tabulated below.

Vintage/age
Boiler manufacturer
Subcritical vs. supercritical boiler
Cyclone boiler fuel firing system vs. other
Once-through boiler circulation vs. other
Boiler draft (balanced vs. pressurized vs. converted)
Pressurized draft vs. other
Steam turbine manufacturer
Size of unit
Boiler reheat (double vs. single vs. none)
Double reheat vs. other
Generator manufacturer
Type of condenser cooling water
Precipitator vs. no precipitator
Type of precipitator
Electrostatic precipitator manufacturer
Mechanical precipitator manufacturer
Ratio of capacity to steam turbine nameplate rating
Reserve shutdown hours
Actual primary fuel

The analysis procedure described above was followed for each of the subject units. A peer group consisting of similar units in North America was formed for each subject unit upon completion of the analysis procedure. The homogeneity of the group for each performance indicator ($\hat{EFOR}$, $\hat{SOF}$, and $\hat{EAF}$) was analyzed. The purpose of the final analysis was to provide statistical information, data distribution, and graphic data for each performance indicator.

To compare the performance of a test unit with that of its peer group, the distribution of the performance indices for the group was presented in a graph. The performance index of the test unit was marked on the graph to indicate its position in comparison with that of its peer group. An example of the $\hat{EFOR}$ distribution obtained using this method are shown in Figure B.2. Similar comparisons can be made for $\hat{SOF}$, $\hat{EAF}$, and other performance indices.

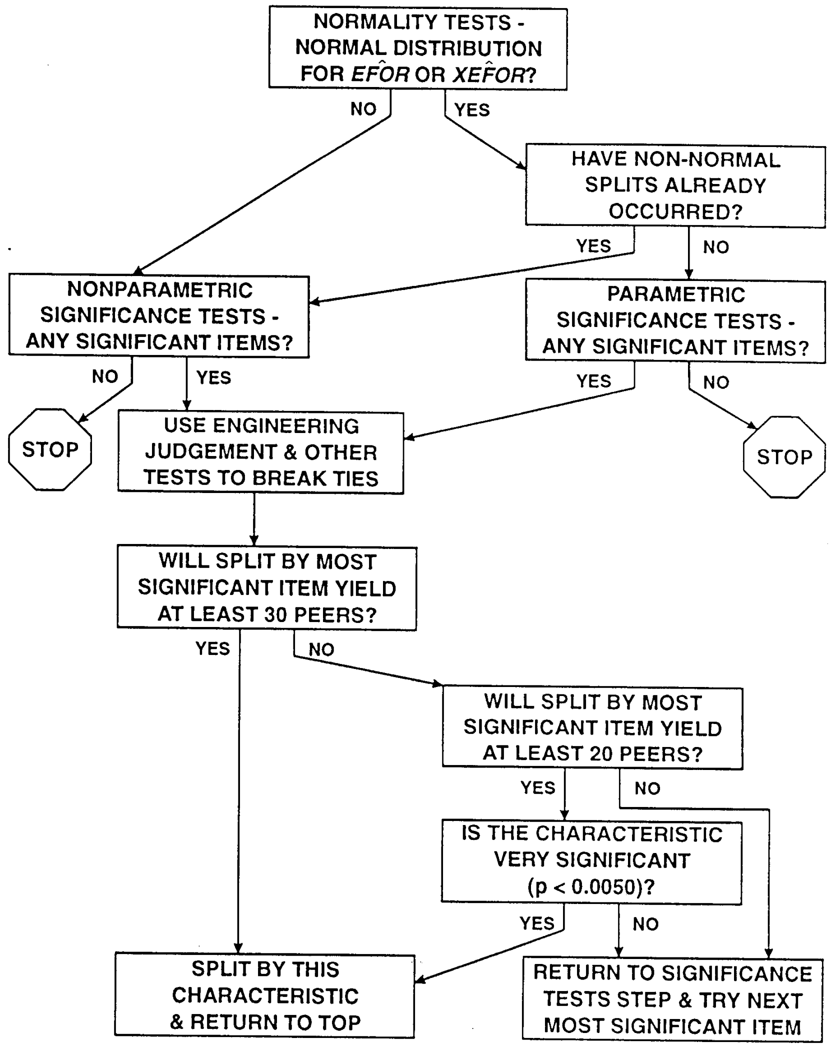
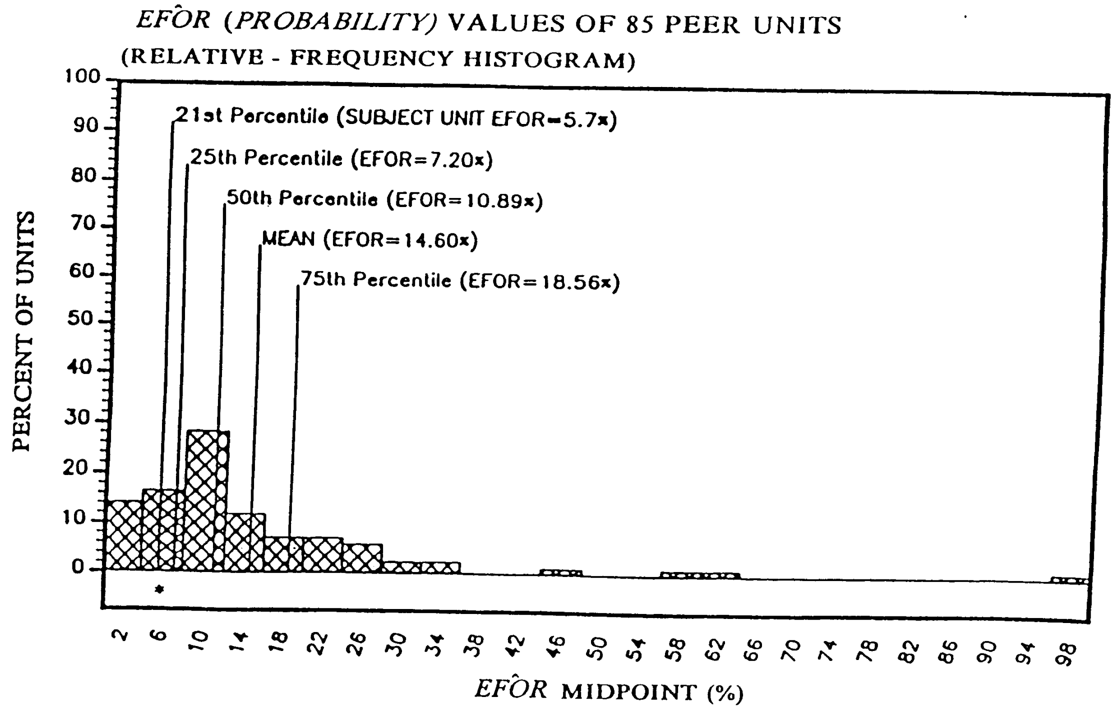


Figure B.1
Flowchart for Analysis of Characteristics for Unit Grouping

TEST UNIT #1



* EFÔR OF SUBJECT UNIT.

Figure B.2
Comparison of Performance Index of a Test Unit with that of its Peer Group

APPENDIX C

Mean and Variance of $\hat{AF}_i$

Let $X(t)$ be a random (or stochastic) process that is 1 if $X(t)$ is available and 0 if $X(t)$ is not available. Partial outages or deratings are not considered in this elementary model. $X(t)$ is then modeled by the usual Markov failure and repair model where λ is the constant failure rate and μ is the constant repair rate.

It is well known that

$$P[X(t)=1|X(0)=1]=\frac{\mu}{\mu+\lambda}+\frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t}$$

and

$$\lim_{t\to\infty} P[X(t)=1]=\frac{\mu}{\mu+\lambda} \tag{C-1}$$

Furthermore, if the process is stationary, i.e.

$$P[X(0)=1]=\frac{\mu}{\mu+\lambda}, \text{ then } E[X(t)]=\frac{\mu}{\mu+\lambda} \tag{C-2}$$

In light of both (C-1) and (C-2) then the availability factor (AF) is defined by

$$AF \triangleq \frac{\mu}{\mu+\lambda}.$$

We seek to estimate AF by $\hat{AF}$ where $\hat{AF}$ is defined by

$$\hat{AF}=\frac{1}{T}\int_0^T X_s(t)dt \tag{C-3}$$

where T is the time period of observation, $X_s(t)$ is a sample function (one generating unit) of the random process $X(t)$.

It is then easy to show that

$$E[A\hat{F}] = \frac{\mu}{\mu+\lambda} = AF$$

(C-4)

That is, $A\hat{F}$ is an unbiased estimator of AF.

The variance of $A\hat{F}$ is given by

$$\begin{aligned} Var[A\hat{F}] &= E\left\{\frac{1}{T^2}\left[\int_0^T \int_0^T [X(t_1)-\frac{\mu}{\mu+\lambda}][X(t_2)-\frac{\mu}{\mu+\lambda}]dt_1 dt_2\right]\right\} \\ &= \frac{1}{T^2}\int_{-T}^{T}\left[1-\frac{|t|}{T}\right]\frac{\mu\lambda}{(\mu+\lambda)^2}e^{-(\mu+\lambda)t}dt \\ &= \frac{2\mu\lambda}{T(\mu+\lambda)^3}\left[1-\frac{1}{T(\mu+\lambda)}(1-e^{-(\mu+\lambda)T})\right] \end{aligned}$$

(C-5)

If $T(\mu+\lambda) \gg 1$ (typical values of $\mu+\lambda$ are of the order of 1 week^{-1}), then

$$Var[A\hat{F}] \approx \frac{2\mu\lambda}{(\mu+\lambda)^3}\frac{1}{T}$$

(C-6)

APPENDIX D

Mean and Variance of $\hat{CF}_i$

Let

CF_i	=	Capacity Factor
$Y_i(t)$	=	MW output of generator i at time t
$E[Y_i(t)]$	=	E [Output | Available] $A(t)$
		where $A(t)$ is the availability at time t
	Call	E [Output | Available] $\triangleq$ $EO_i(t)$

Note that expected output is a function of load, unit commitment strategy, economic dispatch as well as unit capacity, C. Assuming $EO_i(t)=C_iK(t)$, that is, given that the unit is available, the output is proportional to C_i (unit capacity). Then,

$$E[Y_i(t)]=C_iK(t)A(t)$$

and

$$\hat{CF}_i=\frac{1}{C_iT}\int_0^T Y_i(t)dt$$

is unbiased if $A(t)$ is stationary (as assumed) and if

$$E\{\frac{1}{T}\int_0^T K(t)dt\} \triangleq \overline{K}$$

is such that $CF=A\overline{K}$.

Assuming $K(t)=\overline{K}$ then it can be shown that Var $\{\hat{CF}_i\}$ is proportional to $\frac{1}{T}$ where T is the time in the period.

APPENDIX E

Glossary of Statistical Terms

1. **Bias** (n), **Biased** or **Biassed** (adj)

If the expected value, $E\{g(x)\}$, of a statistic obtained from a random sampling, on the ***R. V.(x)*** with the analytical transformation $g(\ )$, is not equal to the parameter or quantity being estimated, the statistic is **biased**.

1A. **Unbiased** or Unbiassed (adj)

Converse of Biased; if the expected value, $E\{\ \}$, of a statistic obtained from a random sampling is equal to the parameter or quantity being estimated, then the statistic is **unbiased**.

2. **Best Linear Unbiased Estimator (BLUE)** (n)

A linear estimator $\hat{\Theta}^*$ is called a **best linear unbiased estimator (BLUE)** for a parameter Θ if it is unbiased and has minimum variance among linear unbiased estimators.

2A. **Best Linear Invariant Estimator (BLIE)** (n)

A linear estimator $\hat{\Theta}^{**}$ is called a **best linear invariant estimator (BLIE)** for a parameter Θ if it has minimum mean squared error, $E\{(\hat{\Theta}^{**} - \Theta)^2\}$, among linear estimators.

3. **Estimator** (n)

A function of observations (or measurements) whose value is used as the (point) estimate of a parameter.

4. **Estimate** (n)

The value of an estimator when the observations have specific numerical values.

5. **Grouping** (gerund)

Grouping is the process of identifying, in a data base, a set of (generating) units which meet specified criteria. The usual purpose of grouping units is to assemble data on generating units which are homogeneous considering one or more characteristics such as size, vintage, design, etc. for improved estimation of a common parameter. However, in some applications, all generating units in a utility may be grouped for the purpose of estimating a performance index for the utility generating system even if the units are not homogeneous.

6. **Homogeneity** (n)

a) K populations are **homogeneous** if the distribution functions are identical.

b) K samples are **homogeneous** (in variance) if the individual sample variances satisfy the **homogeneity** (equality) hypothesis $\sigma_1 = \sigma_2 = \ldots = \sigma_K$ against the alternative that some $\sigma_i \neq \sigma_j$.

7. **Outlier** (n)

a) An **outlier** is a data value that is **far** from the rest of the sample.

b) An **outlier** in a set of data is an observation (or subset of observations) which appears to be **inconsistent** with the remainder of the set of data.

c) An **outlier** is an observation (or subset of observations) which appears to be inconsistent with ('deviates markedly from') the remainder of the sample in which it resides.

8. **Pooled** (adj) (Estimator of a Common Parameter Value)

Suppose that $\hat{\Theta}_1^*, \cdots, \hat{\Theta}_K^*$ are statistically independent BLUEs of Θ. Also, suppose that their variances can be expressed in terms of a **common**, (unknown) scale parameter σ as

$$Var(\hat{\Theta}_1^*) = D_1 \sigma^2, \cdots, Var(\hat{\Theta}_K^*) = D_K \sigma^2,$$

where the factors D_K are known. Then the pooled BLUE for Θ is:

$$\hat{\Theta}^* = D[(\hat{\Theta}_1^* / D_1) + \dots + (\hat{\Theta}_K^* / D_K)],$$

where

$$D = 1/[(1/D_1) + \dots + (1/D_K)], \text{ and } Var(\hat{\Theta}^*) = D\sigma^2$$

Note: Before using such a pooled estimator, one should check that the samples are consistent with the assumption of a common Θ value.

8A. **Pooling** (gerund)

Pooling is the process of aggregating data sets with known common properties to form improved estimates (BLUE) of a common parameter.

Definition Source Texts:

1. G. James and R.C. James, *Mathematics Dictionary,* D. Van Nostrand Co., Inc., Princeton, N.J., 1959.

2. W. Nelson, *Applied Life Data Analysis,* John Wiley and Sons, N.Y., 1982.

3. T. Wright, *Statistical Methods and the Improvement of Data Quality,* Academic Press, Inc., Orlando, 1983.

4. G.J. Hahn and S.S. Shapiro, *Statistical Models in Engineering,* John Wiley and Sons, N.Y., 1967.

5. J.E. Freund and R.E. Walpole, *Mathematical Statistics,* 4th Edition, Prentice-Hall, Englewood Cliffs, N.J., 1987.