United States
Environmental Protection
Agency

Industrial Environmental Research
Laboratory
Research Triangle Park NC 27711

Research and Development

EPA-600/S7-83-055 Feb. 1984

EPA

Project Summary

Variability and Correlation in Raw and Clean Coal: Measurement and Analysis

B. Cheng, K. Crumrine, A. Gleit, A. Jung, D. Sargent, and B. Woodcock

The ability of a coal to comply with an emission regulation depends on a statistical appraisal of the coal sulfur content and heat content determined by random spot or composite sample. Previous studies of the ability of coal to comply with emission regulations have been hampered by inadequacies in the coal sampling method sets which can be used to statistically characterize the variations in coal properties. In this project coal samples were collected at ½- or 1-hour intervals at the inlet to, and outlet from, two coal preparation plants (R&F Coal Company and Republic Steel Corporation). Coal samples from the plants were analyzed for total sulfur, pyrite sulfur, heating value, ash content, and moisture. Values for the organic sulfur and SO_2 emission parameter ($lbSO_2/10^6Btu$) were calculated with ASTM or equivalent procedures. The sample data were evaluated statistically to determine the mean value, variance, relative standard deviation (standard deviation ÷ mean), correlation structure, and skewness. The correlation structure was evaluated by time-series and geostatistic techniques. Time-series techniques proved the most useful.

From this the coal cleaning processes at the R&F and Republic plants reduced the mean SO_2 emission parameter by about 23 and 63 percent, respectively. The relative standard deviation (RSD) of the SO_2 emission parameters were reduced by 26 and 44 percent, respectively. Differences in the reductions in mean and RSD values between plants resulted primarily from differences in the raw coal properties.

For much of the data acquired in this study, strong autocorrelation was indicated. The 30-minute increment data were more highly autocorrelated than composite data collected over longer time intervals.

Currently used time-series models were used to estimate the average number of emission violations produced by a power plant burning the R&F coal (either raw or cleaned) using old procedure. For an emission limit 2.5 standard deviations greater than the mean, and for a 24-hour averaging time, the time series models predicted 12 violations per year for raw coal and 49 for clean coal. Under these circumstances the expected number of violations for clean coal is higher than that for raw coal because of the stronger reliance on autocorrelations used in the clean coal model. These violation frequencies are much greater than the two violations per year that would be predicted from (erroneous) assumption of serially independent coal data. These model results show the importance of considering the effects of autocorrelations when estimating the potential for emission limit exceedance with raw or cleaned coal.

This Project Summary was developed by EPA's Industrial Environmental Research Laboratory, Research Triangle Park, NC, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

Previous studies, to determine if a coal complied with SO_2 emission limits, were based on Gaussian statistics that assume an independently varying sulfur and energy content in coal. Accordingly, the time sequence of the data was ignored, and a distribution curve was used to determine the mean coal sulfur value that ensures compliance. Recent studies indicate that this method may be inappropriate for handling coal data.

These recent studies, however, were hampered because the data sets did not form logically consistent or homogeneous populations sufficient for rigorous statistical analysis. Indeed, much of the reported data were mixed with respect to location, mining method, cleaning method, sampling frequency and procedure, methods of compositing or averaging, definition of a "lot" of coal and the nominal lot size, and analytical laboratory precision. The data sets consisted of coals from different regions, seams, and mines, with inherent geological and engineering differences. Furthermore, the data sets did not necessarily represent their respective regions or seams.

Prior studies were also hampered by the commercial practice of compositing samples and reporting data for relatively large quantities of coal. The relative paucity of data for short time increments (i.e., small coal quantities) has made it difficult to observe and analyze the components of coal variability.

Because of these inherent comparability problems in studying available data that were originally acquired for quite different objectives, prior studies led to only a rough picture of coal sulfur variability. This study, featuring intensive sampling and analysis at selected coal cleaning plants, was designed to overcome the data deficiencies of prior studies.

Objectives of the Study

This program was a controlled experimental study to accurately collect and analyze representative samples of raw and clean coal. By using the same sampling procedures, sample preparation procedures, and laboratory analysis methods for both raw and clean coal samples, representative composite samples of different size lots provided a measure of correlation and variability in the coal sources tested and of the attenuation of sulfur variability from coal preparation plants.

The study had the following major objectives:

1. To measure and evaluate the reduction of variability of sulfur and heating value of raw versus clean coal by collecting and analyzing samples of raw and clean coal at two commercial plants.
2. To measure and quantify the observed variance of sulfur, heating value, and ash associated with the day-to-day changes in raw and clean coal and the variance associated with sampling, sample preparation, compositing, and analysis of raw and clean coal; i.e., measurement uncertainty.
3. To evaluate the relationship between serial correlation and variability by determining if the measured parameters in sequential coal samples are random and, if they are not, to separate the variability of the parameters into correlated and random components.
4. To determine the relationship between lot size and variability.

General Characteristics of Coal Data

Because of the nature of coal formation, there is some structure (as opposed to complete randomness) in the properties of a coal deposit. Superimposed on this structure is a certain amount of randomness. Thus, if samples are taken from a deposit, the statistical treatment of these data should recognize both structural and random characteristics of the samples.

Mining transforms the spatial characteristics of coal into a time sequence of varying coal properties. Alternative mining approaches or schemes within the same deposit provide different time sequences of potential sulfur emissions. These sequences are subsequently modified by coal preparation, coal transportation, coal blending and sulfur emission control. Moreover, the sequence for one coal data set cannot be assumed to be applicable to other coals.

Coal fired utility plants burn raw (unwashed) coal, clean (washed) coal, or a blend of the two. Variability in sulfur content of these coals can produce variability in the level of emissions from the utility stack. The components of variability in coal relate to (1) variability within any one source of coal, (2) variability between the sources of coal being studied, (3) variability associated with sample collection and laboratory analysis, and (4) variability with different size lots of coal.

The autocorrelation properties of coal data are important and must also be evaluated. Ignoring them may result in incorrect statistical models that lead to errors in decisions on the ability of a coal to comply with SO_2 emissions regulations.

Important Statistical Properties

Three statistical measures are important in evaluating coal sulfur data: (1) mean value; (2) variance, a measure of the data spread; and (3) autocovariance or autocorrelation, the statistical relationship of data points to other data points in the same time (or space) series.

The usefulness of autocovariance or autocorrelation is in forecasting future data from past data. Small autocorrelations may indicate that the correlations between past and present values are small, and so past data provide little useful information in predicting future events. If the autocorrelation is large, it is essential to determine the time dependent effect of events through modeling in order to predict future trends. Once these trends have been forecast, the variance can be used to estimate their accuracy.

Experimental Approach and Procedures

Raw and clean coal samples were collected from the R&F Coal Company and Republic Steel Corporation coal preparation plants so that hour-to-hour and day-to-day changes in the coal characteristics could be monitored. These samples were collected and analyzed consistently during the study, using standard techniques so that the variance associated with sampling and analysis would not mask the coal characteristics. The data sets produced were of suitable size to allow application of Gaussian statistics, geostatistics, and time series analysis.

At the R&F plant, the feed and product conveyors were equipped with automatic sampling systems. The raw coal primary sampler took periodic cross-stream cuts of the 4-in.* x 0 feed material, reduced it to 28 mesh x 0 in a hammermill crusher, and split it again through the secondary sampler. The sample collection vessel was left in place for 8 production hours to collect composites; the vessel was removed every 30 minutes during intensive study periods. The clean coal sampling system was similar to the feed system except the clean coal being sampled had a topsize of 2 in. x 0. Both primary samplers were

*EPA policy is to use metric units; however, nonmetric units are used here for convenience. Readers more familiar with the metric system may use the conversion factors at the back.

programmed to take cross-sectional cuts every 6-7 minutes.

The sampling scheme used at the R&F plant was designed to provide as much information on the characteristics of varying lot sizes as possible. Samples were collected at 30-minute increments during two intensive efforts for 22 and 40 hours, respectively. Concurrently, 8-hour composites were collected during the entire study period. The sampling frequencies and time composites represent clean coal lot sizes based on an average 660 tons/hour production rate. Therefore, a 30-minute increment sample represents a 330 ton lot, an 8-hour composite represents 5, 280 tons of clean coal, and a weekly composite is equivalent to a 48,000 ton lot size.

At the Republic plant, the feed and product conveyors were equipped only with primary sampling devices. Unlike the R&F system, the samplers at Republic consisted of a cutter which emptied the material into a hopper for acquisition. The sampler was activated manually through a relay control mechanism. The size of the sample was proportional to the amount of coal on the belt. Normally, the cross-stream cuts produced 200 lb of sample material. No crushing took place during the cleaning operation; therefore, the topsize of both the feed and product coals was 1-½ in. x 0.

The sample scheme used at the Republic plant was designed to provide as many samples or data points as possible during a production day. Samples were collected hourly during the first and fourth weeks, and on the half-hour during the second and third weeks. The data can therefore be compiled to represent hourly samples for a continuous 4-week period and half-hour samples for a continuous 2-week span.

Based on the average clean coal production rate, hourly samples represent 458 tons of coal, a daily composite (6.4 hr) represents 2,932 tons, and a weekly composite is equivalent to 14,663 tons of clean coal.

Results

For much of the data acquired in this study, strong autocorrelation was indicated. The 30-minute increment data from the R&F plant were more highly autocorrelated than composite data over longer time intervals. The data from the Republic plant exhibited weaker autocorrelation than the R&F data. However, results from both plants confirm that serial correlation of coal data does exist over short time intervals.

Two analytical techniques were used to quantify the correlated and random components of the variability in coal data: geostatistics and time-series analysis. Time-series models can be used predictively to generate data sets much longer than the empirical (measured) data set. The random component in the predictive model is obtained from a random number generator. Since a time series model is probabilistic, many different time series, equally likely, may be generated, all based on the same mean, same variance, and same correlation structure. From many time series based on models for one raw and one clean coal data set from the R&F plant, the average number of emission violations by a power plant burning this coal (either raw or cleaned) was determined.

For an emission limit 2.5 standard deviations greater than the mean, and for a 24-hour averaging time, 12 violations per year were predicted for raw coal and 49 for clean coal. These violation frequencies are much greater than the two violations per year predicted from the (erroneous) assumption of serially independent coal data. It should be emphasized that the prediction of violations for the time-series or Gaussian model is derived from the same mean and variance values for the coal.

Histograms were constructed for the raw and clean coals, based on the measured R&F coal data and using the time-series predictive model (see Figures 1 and 2). Each figure shows two emission limits: (1) the limit considered achievable for this coal (2.5 standard deviations above the mean), assuming that the value of each data point is not dependent on the value of any other data point as indicated on the graphs as the Gaussian Emission Limit; and (2) the limit that must be set to ensure an actual average of two violations per year, labelled the Cutoff Emission Limit. These histograms are useful in relating the number of expected violations per year for this coal, for any emission limitation. The histograms based on time-series generation are broader, with higher tails, than corresponding Gaussian curves, demonstrating that many more violations are expected with actual coal than under the misapplication of data independence.

Sampling of both feed and product coals from each of two coal preparation plants, under carefully controlled conditions, has confirmed the results of prior studies. These results indicate that both the mean total sulfur content and the mean emission parameter (lb SO_2 per million Btu) are significantly reduced by the cleaning process, as shown in Table 1.

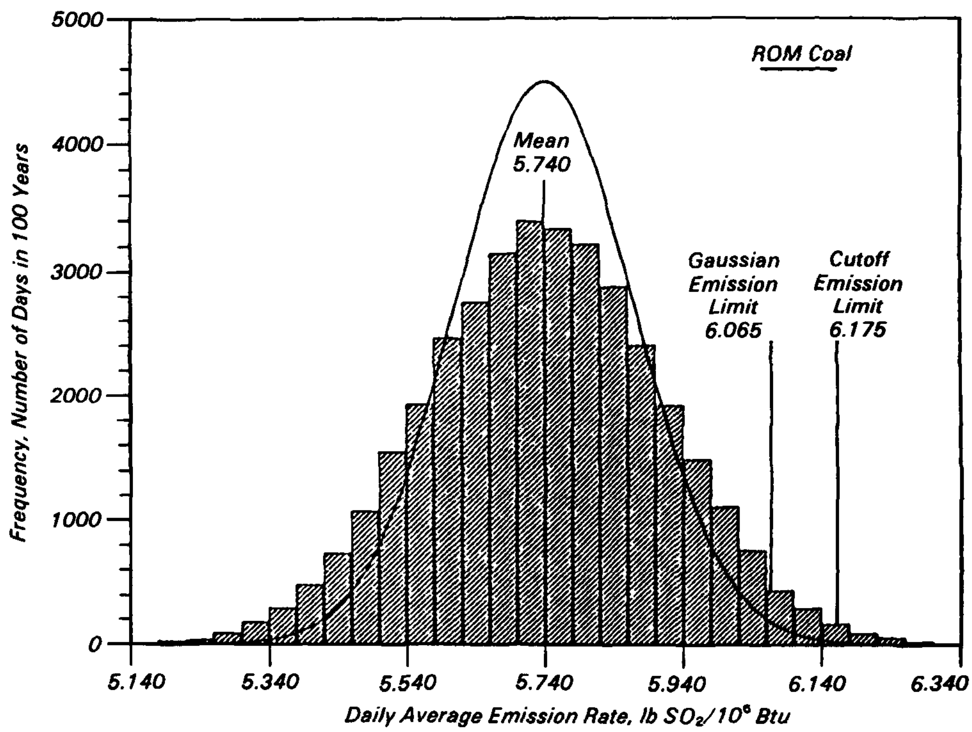

Figure 1. R&F Coal Company—histogram of daily average emission rate based on 100-year time-series generation at 30 minute increments, ROM coal.

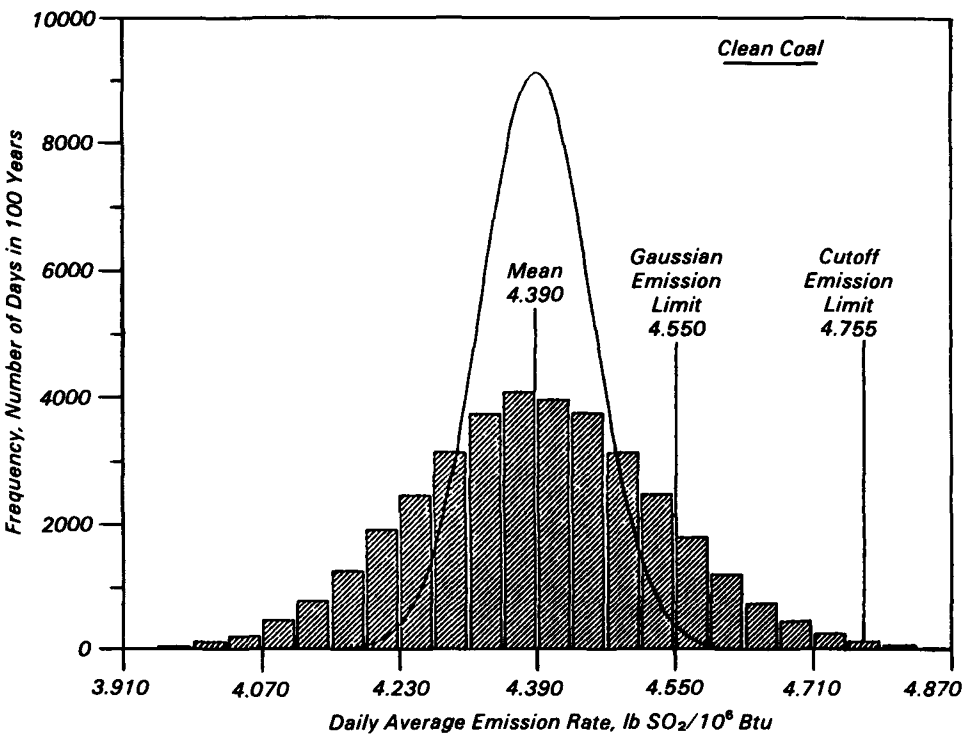

Figure 2. *R&F Coal Company—histogram of daily average emission rate based on 100-year time-series generation at 30 minute increments, clean coal.*

Table 1. *Sampling Results*

		R&F Plant 30-min Increments	Republic Plant 1-hour Increments
Total Sulfur, percent	Raw Coal	3.076	2.576
	Cleaned Coal	2.612	1.309
	Reduction	15.1%	49.2%
Emissions lb SO_2 / 10^6 Btu	Raw Coal	5.476	5.117
	Cleaned Coal	4.237	1.875
	Reduction	22.6%	63.4%

The extent of the reduction is quite different for the two plants. In fact, the 63.4 percent SO_2 emission reduction at the Republic plant is uncharacteristically high for most commercial coal preparation plants. This large reduction in potential sulfur emission results primarily from the washability characteristics of lower Kittanning coal and the operating conditions of a preparation plant processing coal to metallurgical grade specifications. A wide range of reductions between different coal types and preparation plants is consistent with prior findings.

Also confirming the results of prior studies, this investigation documented significant reductions, attributable to the coal preparation process, of the variability in total sulfur and in the emission parameter, as shown in Table 2.

Prior to analyzing the variability in coal data, the measurement uncertainty in the data was independently determined. This uncertainty, attributable to the process of sampling, compositing, sample preparation, and laboratory analysis, provides a quantitative limitation to subsequent explanations of coal variability. All values for aggregate measurement uncertainty were significantly less than the total variations. Real variability in coal characteristics therefore was observed, over and above the measurement noise level.

The time-series predictive model was also used to develop the effect of lot size on variability. The data generated by the time series were mathematically composited into successively longer time intervals (corresponding to successively larger quantities of coal in each interval). The effects of compositing may be expressed either in terms of the averaging time or number of data points (reference lots). For teh R&F plant, a 30-minute clean coal averaging time (a single sample increment) corresponds to a reference lot size of 330 tons. The sample mean variance decreases with increasing lot size, but at a smaller rate than would be expected from serially independent data. This relationship was more pronounced for clean coal than for ROM coal at the R&F plant (see Figures 3 and 4).

Conclusions and Recommendations

Serial dependence (also called autocorrelation) of coal characteristics must be incorporated into any analysis of the ability of either raw or clean coals to comply with SO_2 emission regulations. The misapplication of Gaussian statistics, which assumes serial independence of coal data, leads to a gross underestimation of the frequency of short-term emission violations. time series analysis, which combines serial dependence with a stochastic component to construct a predictive model, provides an alternative to Gaussian statistics. The techniques and computer programs for applying time-series analysis are generally available for use.

Although the two diverse coals studied in detail both exhibited autocorrelation, the magnitude of the autocorrelation component of the total variance differed from one coal to another and from raw to cleaned coal. Therefore, each coal's ability to meet short-term emission regulations must be determined separately until the number of different coals characterized is sufficient to generalize the variability of coal characteristics.

Based on results from the two coal preparation plants studied, one can expect the serial dependence of a coal to be adequately characterized by analysis of consecutive samples, each representing a 30- or 60-minute time increment, if each primary sample is a full-stream cut obtained by an automatic sampler. Extension to 8 hours of the time interval between analyses does not appear acceptable, since the coal properties apparently have a shorter time lag of autocorrelation. Therefore, the conventional sample collection frequencies recommended by ASTM appear to be inadequate to characterize coal variability for short-term emission compliance.

The duration of an extensive sampling and analysis study to characterize the variability of a specific coal should be

Table 2. *Data Analysis*

Measure	Parameter	R&F Plant (30-min Increments) Raw Coal	Clean Coal	Percent Reduction	Republic Plant (1-hour Increments) Raw Coal	Clean Coal	Percent Reduction
Variance	Total Sulfur, %	0.194	0.082	57.7	0.101	0.0072	92.9
	lb $SO_2/10^6$Btu	0.559	0.188	66.4	0.419	0.0172	95.9
Relative Standard Deviation	Total Sulfur, %	0.143	0.109	23.8	0.123	0.065	47.2
	lb $SO_2/10^6$ Btu	0.137	0.102	25.5	0.126	0.070	44.4

several days, sufficient to provide about 80 - 100 consecutive data points at a 30- or 60-minute frequency. The integrity of the consecutive data requirement is of utmost importance in characterizing autocorrelation, and requires that all efforts be directed to avoiding data gaps. To verify the results and to evaluate longer term effects, the intensive study of several days should be repeated at regular intervals over a much longer time span (e.g., 6-12 months).

For this report, intensive testing was conducted at only two preparation plants. The results obtained should be verified for coals other than the ones intensively studied. In particular, cost from other regions and with higher and lower mean sulfur contents should be investigated. The additional studies should address coal feeds to power plants in order to characterize the time variation of coals responsible for boiler emissions. Those studies should include the effects upon variability of blending, stockpiling, and pulvering at the power plants.

Conversion Factors

Nonmetric	Multiplied by	Yields Metric
Btu	1055.1	J
in.	2.54	cm
ton	907.2	kg

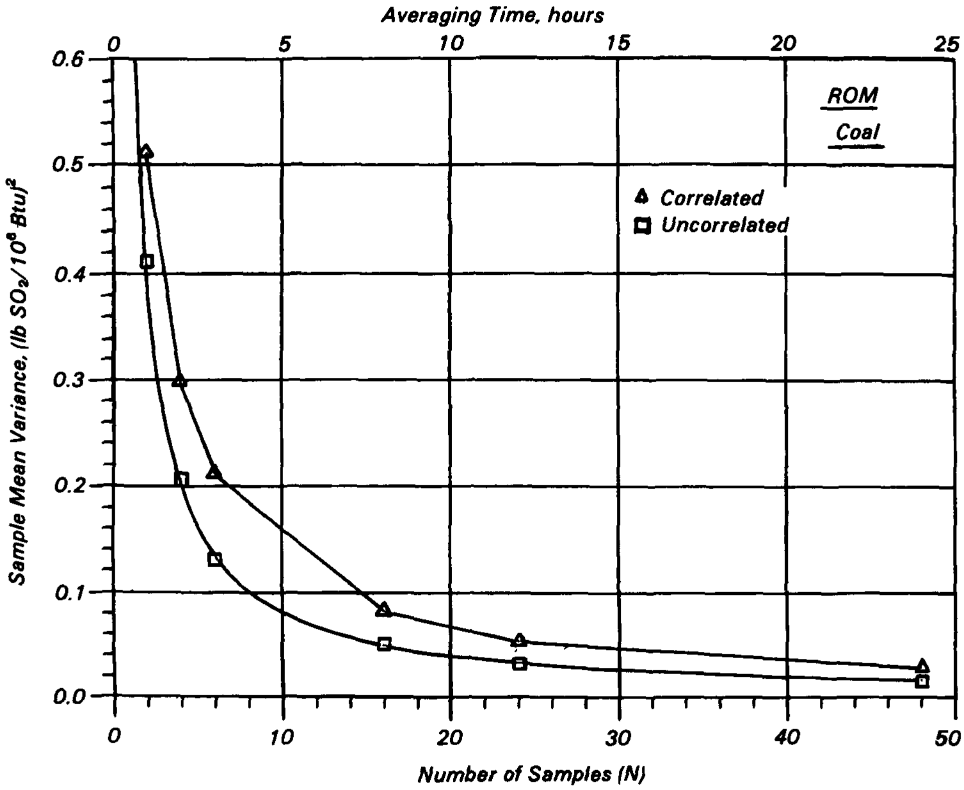

Figure 3. ***R&F Coal Company—sample mean variance of ROM coal as a function of averaging time, uncorrelated vs. correlated data.***

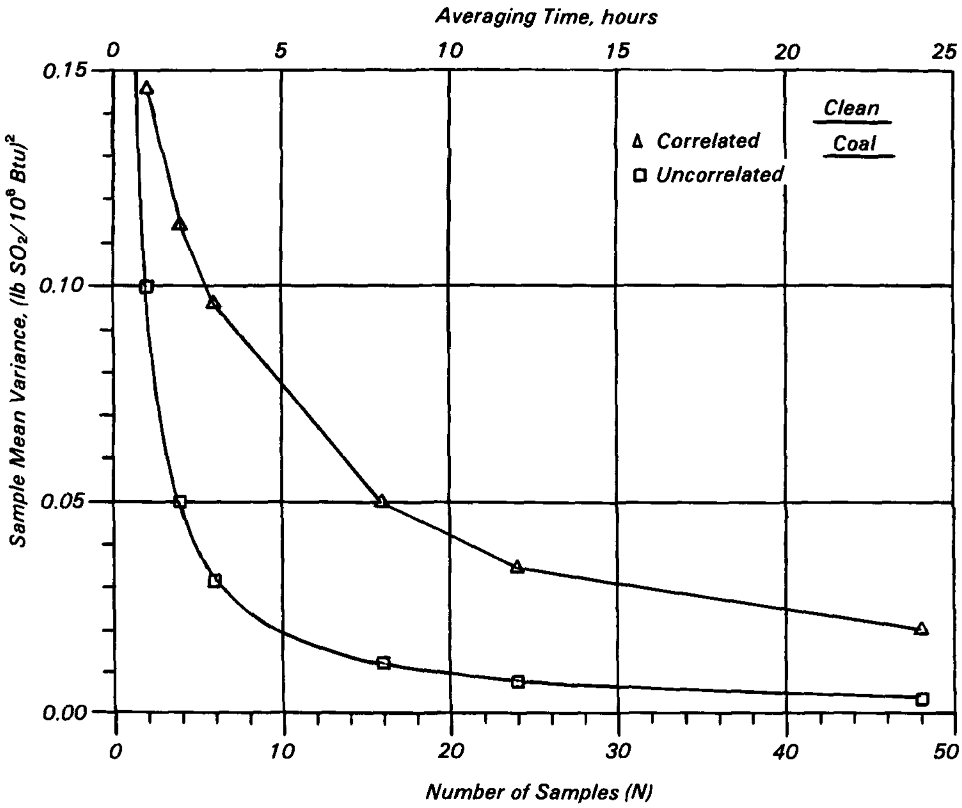

Figure 4. *R&F Coal Company—sample mean variance of clean coal as a function of averaging time, uncorrelated vs. correlated data.*

B. Cheng, K. Crumrine, A. Gleit, A. Jung, D. Sargent, and B. Woodcock are with Versar, Inc., Springfield, VA 22151.
***James D. Kilgroe** is the EPA Project Officer (see below).*
The complete report, entitled "Variability and Correlation in Raw and Clean Coal: Measurement and Analysis," (Order No. PB 84-118 223; Cost: $26.50, subject to change) will be available only from:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-487-4650

The EPA Project Officer can be contacted at:

Industrial Environmental Research Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

United States
Environmental Protection
Agency

Center for Environmental Research
Information
Cincinnati OH 45268

PS 0000329
U S ENVIR PROTECTION AGENCY
REGION 5 LIBRARY
230 S DEARBORN STREET
CHICAGO IL 60604

☆ U.S. GOVERNMENT PRINTING OFFICE: 1984-759-102/851