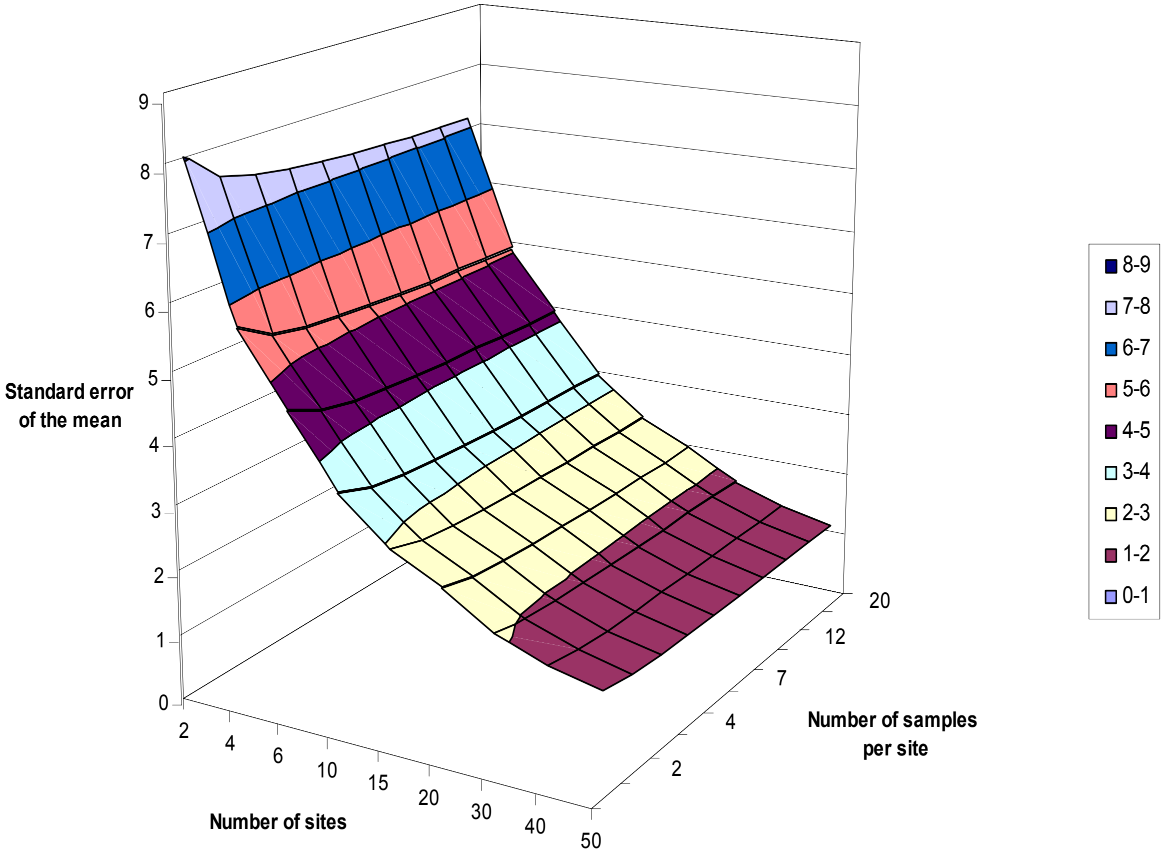
National Decentralized Water Resources
Capacity Development Project



Variability and Reliability of Test Center and Field Data: Definition of Proven Technology From a Regulatory Viewpoint

New England Interstate Water Pollution
Control Commission
Lowell, Massachusetts

September 2005

Variability and Reliability of Test Center and Field Data: Definition of Proven Technology From a Regulatory Viewpoint

Submitted by New England Interstate Water Pollution Control Commission Lowell, Massachusetts

NDWRCDP Project Number: **WU-HT-03-35**

National Decentralized Water Resources Capacity Development Project (NDWRCDP) Research Project

Final Report, **September 2005**

NDWRCDP, Washington University, Campus Box 1150
One Brookings Drive, Cupples 2, Rm. 11, St. Louis, MO 63130-4899

DISCLAIMER

This work was supported by the National Decentralized Water Resources Capacity Development Project (NDWRCDP) with funding provided by the U.S. Environmental Protection Agency through a Cooperative Agreement (EPA No. CR827881-01-0) with Washington University in St. Louis. This report has not been reviewed by the U.S. Environmental Protection Agency. This report has been reviewed by a panel of experts selected by the NDWRCDP. The contents of this report do not necessarily reflect the views and policies of the NDWRCDP, Washington University, or the U.S. Environmental Protection Agency, nor does the mention of trade names or commercial products constitute endorsement or recommendation for use.

CITATIONS

This report was prepared by:

Thomas W. Groves
New England Interstate Water Pollution Control Commission
Boott Mills South, 100 Foot of John Street
Lowell, Massachusetts 01852

New Jersey Department of Environmental Protection
Fred Bowers, Ph.D.

Pennsylvania Department of Environmental Protection
Edward J. Corriveau, P.E.

Independent Contractor
James F. Heltshe, Ph.D.

Northeast Environmental Corporation
John J. Higgins
(formerly Massachusetts Department of Environmental Protection)

Independent Contractor, The On-Site Corporation
Michael T. Hoover, Ph.D.

The final report was edited and produced by ProWrite Inc., Reynoldsburg, OH.

This report is available online at www.ndwrcdp.org. This report is also available through the

National Small Flows Clearinghouse
P.O. Box 6064
Morgantown, WV 26506-6065
Tel: (800) 624-8301
WWCDRE51

This report should be cited in the following manner:

Groves, T. W., F. Bowers, E. Corriveau, J. Higgins, J. Heltshe, and M. Hoover. 2005. *Variability and Reliability of Test Center and Field Data: Definition of Proven Technology From a Regulatory Viewpoint.* Project No. WU-HT-03-35. Prepared for the National Decentralized Water Resources Capacity Development Project, Washington University, St. Louis, MO, by the New England Interstate Water Pollution Control Commission, Lowell, MA.

ACKNOWLEDGEMENTS

Appreciation is extended to the following individuals for assistance in the preparation of this report:

Project Team
Fred Bowers, Ph.D., New Jersey Department of Environmental Protection
Edward J. Corriveau, P.E., Pennsylvania Department of Environmental Protection
James F. Heltshe, Ph.D., Independent Contractor
John J. Higgins, (formerly Massachusetts Department of Environmental Protection)
Northeast Environmental Corporation
Michael T. Hoover, Ph.D., The On-site Corporation, Independent Contractor

New England Interstate Water Pollution Control Commission
Stephen Hochbrunn, Editor
Peter Cavanaugh, Intern

Appreciation is also expressed to the NDWRCDP for their support of this work:

Principal Investigator
Jay R. Turner, D.Sc., Washington University

Project Coordinator
Andrea L. Shephard, Ph.D.

NDWRCDP Project Steering Committee:

Coalition for Alternative Wastewater Treatment
Valerie I. Nelson, Ph.D.

Consortium of Institutes for Decentralized Wastewater Treatment
Ted L. Loudon, Ph.D., P.E.

Electric Power Research Institute
Raymond A. Ehrhard, P.E.
Tom E. Yeager, P.E.

National Onsite Wastewater Recycling Association
Jean Caudill, R.S.

National Rural Electric Cooperative Association
Scott Drake, P.E.

Water Environment Research Foundation
Jeff C. Moeller, P.E.

Members-At-Large:
James F. Kreissl
Richard J. Otis, Ph.D., P.E.
Jerry Stonebridge

Outside Reviewer:
Mark A. Tumeo, Ph.D., P.E.

ABSTRACT

A consortium of environmental agencies concerned with the quality and relationship of test center data to real world data for alternative onsite technologies banded together to formulate this project. The New England Interstate Water Pollution Control Commission (NEIWPCC) is the lead agency for this consortium, which includes the Massachusetts Department of Environmental Protection, the Pennsylvania Department of Environmental Protection, and the New Jersey Department of Environmental Protection. The goals of this research were:

- To develop a statistical and sound scientific relationship between test center data and actual field data of installed alternative technology onsite wastewater treatment systems
- To develop a decision support system to help regulators evaluate the quality and quantity of data submitted for regulatory decisions

While these two major research goals are similar, they require different approaches. The study on these two critical, but different, topics was essential at this stage of technology implementation in the decentralized wastewater field.

State regulatory agencies have been concerned with discrepancies between test center data and real world installations of these technologies. To effectively manage an alternative technology program, state regulatory agencies must be confident in the results that will occur in the real world when constant monitoring, management, and oversight might not be present.

The study included the evaluation of three different alternative onsite technologies that had ample test center and field data sources. Datasets for each system were analyzed statistically using appropriate models. The range of variability was developed, including its variance/covariance structure and the relationships of test center data to predict field system performance. Insights into this variability are beneficial to regulatory staff to optimally design a field-sampling regime, better define a field verification protocol, and better predict expected field performance

The degree of quality assurance/quality control (QA/QC) and management level for the systems was noted, and a baseline was created for comparison of system performance. Relationships between existing data sources—including type, frequency, method of sampling, and quality—were analyzed, developed, and presented.

This research fits into the efforts of the National Onsite Wastewater Recycling Association's (NOWRA) Model Code and the National Decentralized Water Resources Capacity Development Project's (NDWRCDP) goals and objectives to further the beneficial use of proven onsite wastewater technology to solve the nation's small community wastewater needs.

Additionally, a Decision Support System (DSS) tool was developed to help regulators evaluate all sources of data (including test center and field data) to determine the field performance of a technology and guide the regulatory and manufacturing communities on the amount and quality of data needed to accept a technology as “proven.”

The DSS consists of a series of spreadsheets, examples, and documents that guide the user through the ranking of study types, weighting factors, and performance for a stated end-goal. The ideal use of the DSS is as a support tool for regulators making decisions on the evaluation of technology. This is best done by using a multi-reviewer expert panel approach that includes both scientists and regulators, as suggested.

Both goals of this project provide greater insight into where the future of decision-making lies. As the onsite program and industry move toward performance-based codes, this project will improve the baseline understanding of how to assemble, assess, and interpret new and existing datasets to maximize their benefit to the onsite program.

TABLE OF CONTENTS

LIST OF FIGURES

Appendix A

Appendix B

LIST OF TABLES

1 INTRODUCTION

The “standard” or “conventional” septic system, consisting of a septic tank and a gravel leachfield, has served the nation well in the past, but it is now frequently seen as inappropriate or incapable of meeting many of today’s more stringent water quality goals. Onsite wastewater industry professionals, including regulators, designers, and product vendors, are increasingly asked to consider the use of alternative technologies to help accomplish these goals. A problem arises when the industry professionals—particularly regulators—are asked to approve technologies that act as surrogates for the components of the conventional technology systems. The problem is manifested when regulators, or other professionals who control the approval of alternate technologies, seem unable or unwilling to allow vendors and designers to promote these technologies.

There are several reasons why regulators are (or seem to be) recalcitrant and resistant to change. For starters, many states or jurisdictions have laws and prescriptive regulations that restrict decision-making by regulators, making it seemingly impossible for regulators to have the needed flexibility to administer their programs. Most government agencies that regulate onsite wastewater treatment systems have regulations or standards that govern their design. These prescriptions are based on a combination of experimental and empirical science, United States Environmental Protection Agency (US EPA) guidance, and good old-fashioned “experience.” In other words, sometimes the regulations simply express designs for systems that the authorities believe “will work,” and they continue to be used without too much scrutiny because experience suggests that they “do work.” This approach has worked well in the past, but it can lead to problems when someone wants to obtain approval to use an alternative or innovative technology that is not already allowed under the current regulatory framework.

Liability issues can also make onsite regulators resistant to change. For example, if a regulatory agency approves the installation of an innovative/alternative technology that subsequently fails, the regulator and/or the jurisdiction (as opposed to the manufacturer) may be liable for the damage of the failed system, since the regulator approved its use. Even if these liability claims are not valid or cannot be proven, regulators may be reluctant to approve the use of such alternative systems in order to avoid such liability scrutiny.

The problems occur because, in many cases, the rationale for deriving the regulatory requirements is complicated and based on several factors, none of which is clearly expressed in the regulations. They also are not always discussed in the underlying rule proposal documentation. Regulators often admit the process is part politics and part science.

In reality, onsite regulatory strategy is the result of a complex set of factors that can be summarized as follows:

- Social mores and taboos
- Economics
- Consumer protection
- Environmental considerations
- Science

Social Mores and Taboos

The social mores factor can be described as human perceptions, such as aesthetics or taboos, and they partly control how technologies are promulgated. For example, while a sand mound, composting toilet, or spray irrigation system may provide adequate treatment from a scientific point of view, some towns may not want these technologies. People may just not like the looks of a mound, the idea of a composting toilet or outhouse, or the appearance of treated sewage being sprayed on the ground. No amount of science can overcome some of these taboos.

Economics

Economics becomes a factor when technology approvals result in a cost to society or to industry. While a new technology may be proven scientifically to perform better than a standard technology, people may not want to be coerced into using it since it may cost more than a conventional system. It is difficult for regulators to compose a compelling argument for the need to require advanced technology when people are comfortable with conventional technology and do not perceive any personal benefits from a change. In addition, there may be costs to industry when new technology enters the market. While the vendor of the advanced technology benefits, manufacturers of standard technologies may not want to see their market share eroded by the introduction of new products. Thus, issues of fair play, protectionism, and monopolies arise. The manufacturers of conventional technologies may influence decision-making and may challenge proposals to approve new technologies. No amount of science can easily prevail in moving proposals forward when powerful special interests exist.

Consumer Protection

The factor of consumer protection is significant because consumers expect regulators to protect them against system or component failures. Consumers feel they have to blame someone if a technology fails them (or seems to fail them). The manufacturers of onsite technology are not always easy to track down, and even if they were, the vendor would probably—often correctly—argue that the blame should be placed on the designer or homeowner. Warranties are usually limited to the function of a technology, not its performance, and certainly not to its performance in a system with other components that can fail. Home values are affected when systems fail, and homeowners look for someone to blame. Regulators are understandably reluctant to approve new technologies for which they have no experience, and for which they may be blamed if something goes wrong.

Environmental Considerations

Environmental considerations become a factor for decision-making in several circumstances. First, they are an issue when a technology is capable of meeting a performance goal that is adequate in one regulatory jurisdiction but inadequate in another. For example, a technology that can achieve 10 mg/L nitrate 90% of the time may be inadequate in a district where it must meet that standard 100% of the time. A vendor may argue that the performance of that technology is excellent and that it should be approved, but a regulator may be reluctant to do so, since approval may result in complications for a system owner if the water quality goals of the jurisdiction are not met.

Second, it may seem intuitive that a technology, capable of meeting stringent water quality goals, should be approved for use in areas with sensitive environmental receptors. However, these same areas are often preserved from over-development only because standard technology cannot overcome the site limitations and/or achieve the stringent water quality goals. Anti-growth forces view the promulgation of advanced technology as a possible way to develop the “un-developable” lands like wetlands, coastal areas, riparian lands, and hill soils. While the consensus of onsite professionals is that onsite systems should not be used to control growth and land use, one cannot deny that this strategy is commonly used by governments at all levels to control over-development. Regulators are usually unable to influence government decisions made at these levels.

Science

Scientific factors can be summarized as treatment performance, materials quality, and robustness. These factors can all be measured and validated by testing. However, the profession lacks standardized protocols for testing and validation of these factors for many technology types. The National Sanitation Foundation (NSF) in Ann Arbor, Michigan has facilitated standards development, third-party testing, and technology certification for many years; but standards have only been developed for limited types of technologies and components. The NSF Standard 40 tests residential wastewater treatment systems (such as the technologies discussed in this report) for parameters such as Biochemical Oxygen Demand (BOD) or Carbonaceous Biochemical Oxygen Demand (CBOD), and Total Suspended Solids (TSS). As a result, regulators and other industry professionals are compelled to “make up” on-the-spot protocols in order to respond to their immediate needs. The profession needs to establish standard protocols and procedures for testing and validating scientific hypotheses regarding all types of new technologies. The present state of industry science and associated literature often provides less than a complete set of guidance for regulators to use.

Clearly, the onsite system and components “decision-making process” consists of a blend of science, land use, politics, and economics. Except for the scientific factors, regulators have little control over the decision-making process.

The purpose of this project is to provide assistance to industry, regulators, and professionals who use science to make decisions regarding approval of advanced technologies. The assistance consists of two major elements:

- The relative value of field and testing center data, and procedures for professionals to collect adequate field data capable of verifying the performance of technology
- The scientific weight of evidence and how to use it in decision-making through a DSS model, which can be used to validate the relative importance of data from any source

Using one or both of these assistance guides, a regulator or other professional is better equipped to overcome the obstacles presented by politics, lack of protocols, or the currently incomplete professional and scientific information sources.

Background

Onsite regulators and regulatory technical review panels across the country are evaluating a growing number of manufacturers' requests for technology approvals. Technical support documentation for product approval submittals from manufacturers ranges from peer-reviewed journal articles with attached third-party research reports, to simple claims that their “system works just like Product X’s system that you already approved,” with little (or no) supporting third-party research. Testing centers and demonstration projects have been, and continue to be, initiated throughout the country without a comprehensive assessment and national consensus regarding the amount and quality of data necessary for decision-making on what constitutes a “proven technology.”

At the same time, many states, provinces, counties, and communities are reshaping their rules into more performance-based approaches to onsite regulation. The growing environmental focus in onsite wastewater is resulting in a shift in rule revisions, which increasingly emphasize treatment over the traditional emphasis on disposal.

The onsite wastewater program arena is rich with many existing data sources, including:

- Test centers
- Testing organizations
- University test facilities
- Vendor sampling
- State/county/local monitoring

However, the assembly of valid quality data into unified sets needed to confirm statistical trends and relationships is lacking. Understanding statistical relationships can:

- Optimize field-testing protocols
- Reduce unnecessary and costly testing
- Help predict field performance levels
- Enable uniform acceptance of new technology by states, counties, and local onsite oversight and implementing agencies.

Scope

On June 4, 1996 the heads of the state environmental agencies in California, Illinois, Massachusetts, New Jersey, Pennsylvania, and New York signed a memorandum of understanding (MOU) to define a process for the reciprocal evaluation, acceptance, and approval of environmental technologies among the six states. According to the six-state reciprocal MOU, the process would enable participating states to consider data, evaluations, verifications, certifications, approvals, and permits from another participating state as if they had been produced in their respective states.

To implement the reciprocity MOU, the six states selected eleven sample technologies for a pilot project evaluation of this process. The sample technologies included at least one technology of particular interest to each state and represented a full range of environmental technologies for pollution prevention, measurement and monitoring treatment, and control and remediation. Through the pilot project, the six states identified common data evaluation, performance testing, and regulatory review protocols for the pilot technologies and defined the most efficient acceptance and approval process for each technology class. The six states have now used the results of the pilot project to develop guidance for technology developers, vendors, users, and other states. These projects, however, did not include review of onsite wastewater technology.

The MOU initiated an effort in December 1999 by New Jersey, Massachusetts, and Pennsylvania to develop a standard protocol for approving innovative and alternative technology for onsite wastewater disposal systems. Since the initiation of that effort in 1999, Illinois and California have considered becoming participants in the process. Furthermore, discussions with state officials identified a need to develop protocols that deal with this issue in a comprehensive manner and can be applied regionally or nationally. This project represents one attempt to develop a “universal” protocol that any state or entity responsible for approving innovative and alternative technology systems can use, to the maximum extent possible, to approve the work done by others. This protocol has been developed to evaluate and verify the claims made by manufacturers of onsite wastewater systems and components that their products work as expected in real world installations in the field.

The onsite program representatives from New Jersey, Massachusetts, and Pennsylvania, through the New England Interstate Water Pollution Control Commission (NEIWPCC) and with the assistance of The On-Site Wastewater Corporation of Cary, North Carolina, developed and submitted a proposal for this project to the National Decentralized Water Resources Capacity Development Project (NDWRCDP) in the spring of 2003. A contract was awarded to NEIWPCC from NDWRCDP through Washington University in St. Louis in July of 2003. NEIWPCC assembled a Project Team of state regulators and university professionals to commence the project. NEIWPCC subsequently contracted with Dr. James Heltshe of the University of Rhode Island to conduct the statistical analysis and modeling for the project. NEIWPCC also contracted with Dr. Michael Hoover of the On-Site Corporation, Inc. to develop the scientific weight-of-evidence approach to data assessment through the DSS for the project.

The two contractors worked closely with NEIWPCC and the regulator Advisory Committee to identify the technologies and data to be analyzed and help frame the DSS. The contactors and the Advisory Committee, known as the Project Team, met periodically from fall 2003 through summer 2004. Meetings were held in conjunction with other national meetings, such as the National Onsite Wastewater Recycling Association (NOWRA), or the State Onsite Regulators Alliance (SORA), or were scheduled at locations central to the attendees (like northern New Jersey, Baltimore, and Philadelphia). The Project Team also held regularly scheduled conference calls to keep abreast of the project.

The first section of this report evaluates statistical relationships between field and test center data for two parameters—BOD and TSS—and for three advanced treatment technologies—Aquapoint's Bioclere system, Bio-Microbic's Fixed Activated Sludge Treatment (FAST) system, and Orenco's AX treatment system. These technologies were selected because of the large quantity of data that was available from both field locations and test centers. Note that the objective of this project was not an evaluation of the technology brands or their pretreatment performance, but instead a determination of statistical relationships for known datasets.

The second section of this report provides a decision support system. The system can be used to integrate the use of multiple sources of data, including test center and field data, to evaluate the field performance of a technology and to guide the regulatory and manufacturing communities on the amount and quality of data needed to accept a technology as "proven."

These two projects could easily have been developed separately. Taken together, however, their complementary relationship provides greater insight into where the future of decision-making lies. As the onsite program and industry moves toward performance-based codes, these projects will improve the baseline understanding of how to assemble, assess and interpret new and existing datasets to maximize their benefit to the onsite program.

Objectives

The primary objectives of this project were:

1. To assemble valid quality test center and field data into unified sets and evaluate their relative qualities

2. To analyze these datasets statistically to determine whether test center and field data distributions are similar or dissimilar

3. To predict field performance relationships, if data distributions are similar

4. To develop a sampling protocol for future approval of residential wastewater system components, if the test and residence datasets are dissimilar

5. To develop a decision support system for ranking or weighting different types of data that guide regulators and manufacturers on the possible combinations of test center and field data needed to allow state/county/local approvals of new technology as "proven."

As stated earlier, the purpose of this project is to provide assistance to industry, regulators, and professionals who use science to make decisions regarding approval of advanced technologies. Objectives 1–4 relate to the first project goal of examining the relative value of field and testing center data, and the fourth objective attempts to present procedures for professionals to collect adequate field data capable of verifying the performance of technology. Objective 5 relates to the second project goal of presenting a proposed model for regulatory decision-making that can be used to validate data from any source, such as the decision support system.

Other key objectives of this project were to:

- Support and contribute to the acceptance of the NOWRA Model Code
- Build capacity and understanding among those involved in the onsite industry (including vendors, testing organizations, state regulators, consultants, implementing and management agencies, and the public)
- Provide a CD Resource Tool on the collection, assembly, analysis, and use (weighting and ranking) of data collected at test centers and in the field that gives regulators confidence in the predictable performance of new onsite technology

.

2 TEST CENTER VERSUS FIELD DATA: A STATISTICAL COMPARISON AND EVALUATION

The committee reviewed commonly known onsite/decentralized pretreatment technologies and decided to use Orenco's Advantex AX, Aquapoint's Bioclere, and Bio-Microbic's Fixed Activated Sludge Treatment (FAST) due to the number of datasets available from both testing sites and field sites. Committee members Fred Bowers, Ed Corriveau, and John Higgins compiled the field and test center data for the three selected technologies from a multitude of sources, including the National Sanitation Foundation (NSF), Environmental Technology Verification (ETV), National Onsite Demonstration Project (NODP), Massachusetts Department of Environmental Protection (Mass DEP), and manufacturers

They screened the data and amassed a database to facilitate statistical processing and analyses. The committee selected the three technologies that had the most intersection points between test center and field data sources. They wanted to ensure a high level of NSF and ETV data sources to statistically compare the data.

The datasets evaluated consisted of many residence sites with single sampling frequencies and a few sites with repeat sampling. This scenario enabled irregular time series observations of datasets for each technology and long, regular time series of observations from NSF and test center datasets. This is similar to a clinical medical or drug trial using actual human subjects tested at irregular intervals over time and a trial using test species in a controlled lab setting that is looked at regularly over time. Consequently, it does not make sense to statistically test for differences between the human field subjects and actual lab test species. Likewise, it does not make sense for us to statistically test for differences between the residence datasets and their corresponding test center datasets. Each dataset was looked at as being objectively distinct unless the statistics showed them to be otherwise similar.

Parameters of each data distribution were estimated (such as resident sites and test center sites) and fit into probability models to each separate dataset. Descriptive statistics were used to illustrate the possible need to transform these datasets for future statistical inferential procedures. Seasonal trends were also looked at to determine if the data needed to be subset by season. However, temperature information was not available for these datasets. Finally, for the residence datasets, sources of variability were estimated and the estimates were used to develop a sampling protocol for future residential system evaluations. The sampling protocols were developed for both "raw reported" and "log transformed" data to determine if a transformation was necessary.

Every effort was made not to be overly judgmental about the quality of the data reported for analysis, mainly because of the lack of available data in large enough datasets.

Residential data was gathered on Biochemical Oxygen Demand (BOD) and Total Suspended Solids (TSS) collected from known types of treatment units serving known facilities with five bedrooms or less (or test center simulations thereof) that preferably experience "true" winter (cold ambient temperature conditions on a regular basis for at least some part of the year).

Once the data were received and known to be residential sites in a seasonally warm/cold climate, data that met these characteristics were added to an Excel spreadsheet. The merged datasets were then only edited for missing observations, "0" values, and duplicate observations on the same sampling date at a site. Missing observations were removed, "0" observations were considered as missing and thus removed. Duplicate observations on the same date at the same site were considered as "split" samples at a site and were averaged but recorded only once at a site/date combination.

No data were discarded based on who collected the sample or provided the data, nor were the data eliminated or interpreted for any other reasons, such that it was not considered whether the range of values was outside of what was normally expected. Quality or validity assumptions were not made about any of the datasets except as mentioned earlier for missing observations, "0" values, and duplicate observations on the same sampling date at a site. This information is referenced in the Quality Assurance Final Report in Appendix C.

While there is a tendency to try to extrapolate test center data to resident site data, each of these two datasets is distinct. Each has great value depending on how anyone wants to value and use the data as explained further in the proposed Decision Support System (DSS). Extrapolated test center data would be less representative of accomplishing the goal of developing a model to predict the long-term performance of systems in the field. There are simply too many variables inherent with how each field system is operated and maintained, and how each system is independently and differently loaded.

For the purposes of this study, it should be noted that operation and maintenance (O&M) records for any of the systems in this study were not collected and analyzed. The authors fully realize the influence of O&M on the long-term performance of these systems. Although O&M records are available for the test center sites, they were not available for the numerous residence sites. Numerous residence data points would have been disregarded if O&M records were required. This report serves as a real world example and comparison of the performance of field sites as they exist and as they are operated and maintained today. One of the goals of this study was to see if there was a basic relationship between residence sites and test center sites, and as stated above, no data were discarded. Some regulatory programs have the ability and the resources to track O&M of these systems, but most do not.

Task I: Data Editing, Organization, and Presentation

The datasets used in these analyses consisted of data from three technologies:

- Orenco's Advantex AX
- Aquapoint's Bioclere
- Bio-Microbic's FAST

The data were a time series of observations collected on two variables: TSS and BOD. The data were collected at resident field sites (residences) and at test centers used for the evaluation of the National Sanitation Foundation's (NSF) Standard 40. Table 2-1 gives the specifics of the datasets with respect to:

- Time periods of collection
- Sample size
- Number of sampling sites
- Proportion of longitudinal data (A longitudinal dataset is defined as a dataset at a site with at least two time series observations.)

Table 2-1
Summary Information for All Technologies and Variables

Technology	Variable	Location	Sample Size	Number of Sites	Proportion Longitudinal	Sampling Dates	
						Start	Finish
Bioclere	BOD	Sites	691	119	103/119	10/27/1993	10/15/2003
		NSF	107	1	1/1	8/30/1999	2/25/2000
	TSS	Sites	676	119	102/119		
		NSF	107	1	1/1		
Advantex	BOD	Sites	95	22	21/22	1/8/2002	7/7/2003
		NSF	108	1	1/1	5/22/2001	11/16/2001
	TSS	Sites	104	22	21/22		
		NSF	107	1	1/1		
FAST	BOD	Sites	2795	529	461/526	12/1/1995	8/16/2003
		NSF	135	1	1/1	1/1/2002	12/31/2002
	TSS	Sites	2775	529	460/526		
		NSF	135	1	1/1		

For the purposes of this study, results are stated for both transformed and untransformed data. Whether a transformation should be applied to the data is not immediately obvious from looking at the histograms of the datasets. The decision to transform or not needs to be made after the data are collected. Ultimately, one is going to select several (*k*) sites and sample (*n*) observations within a site. One will compute site means and average over the *k* sites to get an overall mean. If one samples many sites (*k* large) then the Central Limit Theorem of statistics states that the average will follow a Normal distribution and no transformation of the data is necessary if one tests for mean difference or constructs Confidence Interval estimates of the mean. If *k* is small, then the site means may not be "Normally" distributed and a transformation may be necessary. The transformation may be necessary if there are "outlier" site means or "outlier" observations within a site.

As stated earlier, although not immediately obvious at first from looking at the histograms of the datasets, the histograms of the raw data clearly illustrate that there are outliers in the datasets that were used in the analyses, thus a transformation is needed. The findings present the results for both transformed and untransformed data, but for the purpose of the analysis, a transformation of the data is necessary.

Figures 2.1 (a-d) through 2.3 (a-d) in Appendix A show the frequency distribution and superimpose the fit of a Normal and a Lognormal density function for each dataset. For the FAST technology, 32 out of 2,930 BOD values were greater than 100. This is 1.1% of the total number of observations. Also for the FAST technology, 17 out of 2,910 TSS values were greater than 100. This is 0.6% of the total number of observations. For the Bioclere technology, 17 out of 798 BOD values were greater than 100. This is 2.1% of the total number of observations. Also, for the Bioclere technology, 21 out of 783 TSS values were greater than 100. This is 2.7% of the total number of observations. For the Advantex technology, 15 out of 203 BOD values were greater than 20. This is 7.4% of the total number of observations. There were no values in the dataset greater than 100. Also, for the Advantex technology, 7 out of 211 TSS values were greater than 20. This is 3.3% of the total number of observations. There were no values in the dataset greater than 100.

Figures 2.4 through 2.8 in Appendix A show the distribution of the data by site number for all three technologies using a log (10) scale for TSS or BOD to better show the data. Based upon these figures, potential outlier and extreme values were identified. All data values were kept for statistical analysis and presentation. For some of the sites, there was more than one data point, but it was the same value, thus, it appears as only one point in the figure. This explains the 21/22 in Table 2-1 for the Advantex technology. Figures 2.9 through 2.11 in Appendix A show the frequency distribution of the number of time series samples taken at a site for each variable and technology.

During the data evaluation, a minimum detection limit of the laboratory analysis of TSS and BOD was observed. Figures 2.14.a through 2.14.b in Appendix A show the BOD and TSS values for the FAST dataset plotted against time. For BOD there seems to be a change in the minimal detection limit around December 1999. It appears that the minimal detection limit was about 2 before December 1999 and then the minimal detection limit was increased to 4 after December 1999.

For TSS there appears to be a change in the minimal detection limit around June 2002. For the Advantex technology there does not appear to be any change in the minimal detection limit over time (Figures 2.12.a and 2.12.b). Similarly, for the Bioclere technology there does not appear to be any change in the minimal detection limit over time (Figures 2.13.a and 2.13.b).

Table 2-2 shows the minimum values for each variable for each technology separated by residence site data and NSF test datasets. This discussion is illustrative of the number of potential outliers in the individual datasets. For the purpose of this study, all the data points were kept. In order to deal with potential outliers, the data were analyzed with and without transformation. No outliers were removed. The data was not subset because of a possible change in detection limits that might have occurred in the middle of the collection process. There did not seem to be a need to test whether resident means (before) were equal to resident means (after) the change in detection limits. All the data was pooled and this change in detection limits contributed to the overall variability, just as outliers contributed to the overall variability.

Table 2-2
Minimum Observed Values by Variable and Technology Separated by Site and NSF Datasets (mg/L)

	Bioclere		**Advantex**		**FAST**	
	TSS	BOD	TSS	BOD	TSS	BOD
Residences	1.0	2.0	0.5	0.5	1.0	1.0
NSF	2.0	4.0	2.0	2.0	5.0	5.0

Figures 2.15 through 2.17 in Appendix A show the relationship between BOD and TSS for each technology. Log 10 scales were used to better show the data. These figures, along with Figures 2.4 through 2.8, can be used to identify outliers or extreme values in the different datasets.

Each dataset consists of time series of observations taken at residence sites or the NSF testing site. Since the time series cover several years (Table 2-1), it was necessary to determine if there was any annual trend within the datasets. The residence dataset is an irregular time series for each site. Most sites were visited only two or three times over many months. Thus, a formal time series analysis for seasonal trends could not be done. The test dataset was a “nice” regular spaced time series; however, this time series did not cover the entire year. All observations recorded for an individual month were combined, regardless of the year in which the data were collected.

Figures 2.18 through 2.20 in Appendix A show the monthly means for the site data and the NSF dataset. Figures 2.21 through 2.23 in Appendix A show the monthly medians for the site data and the NSF dataset. Medians are less affected by extremes in the datasets. The 25th percentile and the 75th percentile values are also given to show the variability in the datasets. The variability shown in these mean and the median figures are functions of “between” sites and “within” sites for the residences datasets but only “within” the datasets for the NSF data because there are no replicate NSF sites. Formal statistical analysis of between and within variability is the focus of Task II.

Table 2-3 gives the overall means and medians for each technology, separated by residence sites and NSF test site.

Table 2-3
Means and Medians for Each Technology Separated by Residence Sites and NSF Test Sites (mg/L)

		Means		Medians	
		Residences	NSF	Residences	NSF
Advantex	TSS	7.93	3.99	6.0	3.0
	BOD	11.34	4.63	6.9	3.0
Bioclere	TSS	22.06	5.29	10.0	5.0
	BOD	23.15	11.07	12.0	11.0
FAST	TSS	14.32	7.81	7.2	6.0
	BOD	15.60	9.60	8.9	9.0

In looking at Figures 2.18 through 2.23, it is difficult to see any seasonal trend in the data. Most monthly means are similar in magnitude and there does not appear to be any elevation or reduction in the response during colder months. The analysis of the time series indicated that there was no observable inter-annual trend. Therefore, it did not matter what month the samples were taken. This conclusion is only valid for the parameters that were analyzed for this study (BOD and TSS). It is assumed that other parameters, such as nitrogen, would be more apt to be affected by the time of the year. It needed to be determined whether there was a seasonal pattern that might be a function of temperature. In other words, do the cold months appear different from the warm months? The data at the residence sites were gathered with no time series plan in mind. Thus, the best option was to pool all individual months data and look for trends in the data. The monthly means were greatly affected by outliers, so medians were presented for each month of the year.

The overall means of the residence datasets are all larger, by a factor of approximately two, than the corresponding overall means for the NSF test datasets. Since these differences may be due to a few extremely large values within a particular dataset, the medians were calculated to see if differences in datasets might change. Looking at the medians, which remove the effect of extremely large observations in the datasets, one sees that the medians differ by a factor of two for Advantex TSS and BOD as well as Bioclere TSS. However, for Bioclere BOD and FAST TSS and BOD the medians are similar. Because of the outliers in the datasets, medians are the best and most appropriate statistics to use for the discussion.

Task II: Sources of Variability and Data Comparison

Maximum likelihood estimates of the variance components were estimated for each variable, each technology, residence, and NSF test data using untransformed data and log10 transformed data. Milliken and Johnson (1984) give a detailed description of the different estimation procedures for variance components.

Each outcome variable has a total variability often denoted by εi in a standard Analysis of Variance model:

$$Y_i = \beta_0 + \beta_i + \varepsilon_i$$

i = ith treatment effect

β_0 = overall mean of all observations

β_i = effect associated with the ith treatment level

ε_i = random error associated with the ith treatment level

When there are repeat measurements over time from a random effect, residences (assumed to be randomly selected from an infinite population of possible residences), the model may be rewritten as:

$$Y_{ij} = \beta_0 + \tau_i + \varepsilon_{ij}$$

i = ith residence effect

j = jth repeated observation on ith residence

β_0 = overall mean of all observations

τ_i = random effect associated with the ith residence

ε_{ij} = random error associated with the jth observation on the ith residence

The total variance estimate e for ε_{ij} can be separated into within-residence variability, e_w, and between-residence variability, e_b, where $e_w + e_b = e$ is the total variability of the observed values.

The Analysis of Variance model is used to estimate variance components. The sources of variability were estimated, which were between sites and within sites for the residence dataset and only within for the test dataset. The estimation procedure took into account the fact that the data was serially correlated within an individual residence and within the test dataset.

For the residential data, each site can be considered a random effect and the variability of the outcome is attributed to both between sites variability and within sites variability. For the NSF data there is only one site, so the estimate of total variability can only be attributed to within-sites NSF variance.

Table 2-4 gives the maximum likelihood estimates of the variance component for untransformed data for between sites and within sites. For all technologies and both variables, the between-sites variance exceeded the within-sites variance by a factor of 3 to almost 300. The within-sites variance for the NSF datasets was less than the within sites residences variance for all technologies and both variables with the exception of the Advantex technology for TSS. The within-sites variance for the residences exceeded the within-sites variance for the NSF data by a factor of 2 to 14. For the Bioclere and FAST technologies, there is a large total and between-sites residence variance.

Table 2-4
Variance Component Estimates for Between Sites and Within Sites for TSS and BOD for Three Technologies, Separated by Residences and NSF Data (Untransformed Data)

		Residences			**NSF**
		Within	**Between**	**Total**	**Within**
Advantex	TSS	7.2	49.1	56.3	21.6
	BOD	29.1	100.9	130.0	15.2
Bioclere	TSS	154.9	1946.1	2101.0	5.0
	BOD	272.2	2286.6	2558.8	20.0
FAST	TSS	82.5	2484.3	2566.8	28.2
	BOD	65.0	374.8	439.8	14.5

Table 2-5 gives the maximum likelihood estimates of the variance component for log10 transformed data for between sites and within sites. The clear pattern of between-sites variance being larger than within-sites variance using the untransformed data (Table 2-4) is not as pronounced in Table 2-5. Generally, the relationship exists that the largest source of variability is between residence sites.

Using the log10 transformed data reduces the effect on the variance estimates of extremely large values that show up in the residences datasets. Figures 2.1 through 2.3 show that the lognormal distribution generally fits these residences' datasets better than the normal distribution due to large values in the datasets. This is also true for the NSF datasets. Thus, applying a log10 transformation to the datasets gives better estimates of the variance. Task III uses these variance estimates to develop a sampling protocol for residences sampling. Both untransformed and log10 transformed data estimates of variance are used separately.

Table 2-5
Variance Component Estimates for Between Sites and Within Sites for TSS and BOD for Three Technologies, Separated by Residences and NSF Data (Log10 Transformed Data)

		Residences			NSF
		Within	Between	Total	Within
Advantex	TSS	0.017	0.108	0.125	0.060
	BOD	0.047	0.054	0.101	0.080
Bioclere	TSS	0.071	0.093	0.164	0.033
	BOD	0.057	0.089	0.146	0.031
FAST	TSS	0.051	0.057	0.108	0.037
	BOD	0.038	0.064	0.102	0.025

Table 2-6 gives the Coefficients of Variation (CVs) defined as the standard deviation divided by the overall mean for each dataset. The standard deviations use the square root of the between, within, and total variance estimates.

Table 2-6
Coefficients of Variation (CVs) Estimates for Between Sites and Within Sites for TSS and BOD for Three Technologies, Separated by Residences and NSF Data (Untransformed Data)

		Residences			NSF
		Within	Between	Total	Within
Advantex	TSS	0.34	0.89	0.95	1.16
	BOD	0.48	0.89	1.01	0.84
Bioclere	TSS	0.56	2.00	2.07	0.42
	BOD	0.71	2.07	2.19	0.40
FAST	TSS	0.64	3.49	3.54	0.68
	BOD	0.52	1.24	1.34	0.40

Table 2-7 gives the Coefficients of Variation for the log10 transformed data.

Table 2-7
Coefficients of Variation (CVs) Estimates for Between Sites and Within Sites for TSS and BOD for Three Technologies, Separated by Residences and NSF Data (Log10 Transformed Data)

		Residences			NSF
		Within	**Between**	**Total**	**Within**
Advantex	TSS	0.18	0.45	0.48	0.49
	BOD	0.25	0.26	0.36	0.51
Bioclere	TSS	0.26	0.30	0.39	0.26
	BOD	0.22	0.27	0.35	0.17
FAST	TSS	0.25	0.27	0.37	0.23
	BOD	0.20	0.25	0.32	0.17

The statistical analysis demonstrates that the variability associated with "test center data" was significantly less than the variability of data collected from "real world" situations. Therefore, the two data distributions are dissimilar, and one data distribution set (such as test or field) cannot be used to accurately predict the other.

Task III: Development of a Sampling Protocol

Using the information from Tasks I and II, a sampling protocol was developed for future sampling of technology performance. A look at Figures 2.18 through 2.23 shows that there is no obvious inter-annual trend for the BOD and TSS data used in this study. Consequently, this suggests that future sampling at a residence for TSS and BOD can be started and finished at any time of the year. A minimum 12-month sampling period is suggested. This may not be appropriate for other temperature-affected parameters, such as nitrogen, or for different technologies. To evaluate the variable response of a "new" technology, one should obtain multiple samples at several residence sites. An ideal sampling plan should have k residence sites, each sampled n times over a period of time. The overall mean response for a variable will be the sum of these $k \times n$ observations divided by $k \times n$. The variance of the overall mean is a function of between-sites variance and within-sites variance.

$$\text{Var(mean)} = \text{(Between variance)}/k + \text{(Within variance)}/nk$$

where: n is the number of samples at a residence site and k is the number of residence sites.

The variance between sites exceeded the within-sites variance for all technologies, both variables, and for both untransformed and log10 transformed data (Table 2-4 and Table 2-5).

Thus, a sampling plan with more residence sites and fewer samples at a site will be more efficient in reducing the variance of the overall mean. Using the between-sites and within-sites variance given in Table 2-4 and Table 2-5, the variance of the overall mean can be calculated for different combinations of *k* and *n*. Using the untransformed data estimates of the between-sites and within-sites variances, the variance of the mean is reduced faster by increasing the number of residence sites. It is affected only slightly by increasing the number of samples per residence.

This is shown in Figures 2.24 and 2.25 in Appendix A for the Advantex technology and variable TSS. The surface shown in Figure 2.24 is consistent for different numbers of samples per residence site. Figure 2.25 shows the cross section of the surface illustrating that the standard error of the mean is reduced more by increasing the number of residence sites than by increasing the number of samples per residence site. This pattern is consistent for all technologies, both variables and for both untransformed and log10 transformed.

Appendix B includes the figures for the other combinations of technology and variables for both untransformed and log10 transformed data (Figures B-1–B-22). In all cases, the standard error of the overall mean is reduced more by increasing the number of residence sites. The magnitude of the standard error is a function of the total variance of the variables for each technology. For those technologies and variables with a large difference between the between-sites variance and the within-sites variance, the more efficient sampling protocol is to select more residence sites with few samples per site.

For the log10 transformed data, the effect of increasing the number of residence sites is not as dramatic since the differences between the between-sites variance and the within-sites variance is not as great as for the untransformed data. On the log10 transformed scale the standard error of the overall mean can be reduced if four samples are taken at a residence site. Taking more than four samples results in only a minimal reduction in the standard error. Thus, a sampling protocol seems to be to take four samples at a residence and take as many residences as necessary to achieve the acceptable standard error of the mean.

If an acceptable standard error of the mean can be established, then the figures in Appendix B and Figures 2.24 and 2.25 in Appendix A can be used to develop a sampling protocol that can achieve this standard of error. Different combinations of number of residence site and samples per site give approximately the same standard error of the mean. However, as stated earlier, the most "bang for your buck" in reducing the standard error of the mean can be achieved by selecting more residence sites. For a total number of $m = k \times n$ samples that need to be collected and analyzed, it is more efficient to increase *k* (the number of residence sites) and reduce *n* (the number of samples taken at a site). This is true regardless of the technology or the variable measured, because for all cases the between-residence variance was greater than the within-residence variance.

Figure 2.26 in Appendix A gives an example for the Advantex technology, TSS variable of the sampling requirements in terms of the number of sites and samples per site if one wants 84%, 95%, or 99% confidence in an interval estimate of the overall mean. For this example, the sampling requirements for a one-sided upper confidence limit were determined.

In practice, one might want to determine if the upper confidence limit about the overall mean exceeds some criterion level. Three different *n*'s were selected for number of samples to be taken at a site: *n*=1, 4, and 12 represents yearly, seasonally, and monthly sampling, respectively. The 84%, 95%, and 99% confidence intervals involve multiplying the estimate of the standard error of the overall mean by 1, 1.645, and 2.33, respectively, and adding this result to the overall mean.

Clearly, the confidence interval width increases with the degree of confidence. This is the case when the sampling protocol is fixed for *n* and *k*. If the confidence interval width is fixed, then to achieve this fixed width, the sampling requirements must be changed to achieve the different confidence intervals. Thus, to get the same fixed confidence interval width, the estimate of the standard error of the mean must be reduced by increasing the sample size. An 84% confidence interval width is 1 × (standard error). A 95 % confidence interval width is 1.645 × (standard error). Thus, to achieve a fixed confidence interval width, the standard error must be reduced by a factor of 1.645, that is, divide the standard error by 1.645. For a 99% confidence interval, this factor is 2.33.

In Figure 2.26, a fixed confidence interval width of 3 was selected. Thus, the 84% confidence interval is only 1 × (standard error). Figure 2.26 shows that for *n*=1, about *k*=7 sites are needed. For the 95% and 99% confidence intervals to have the same fixed interval width, one needs about *k*=18 and *k*=37, respectively. For *n*=4 or *n*=12, the number of sites would be reduced only slightly, if at all, since the between-sites variance is the most important factor driving these sampling plans. Similar sampling plans can be determined for all technologies and both variables, TSS and BOD, from Figures 2.24 and 2.25 in Appendix A and the figures in Appendix B using untransformed and log10 transformed data.

The sampling protocol results and confidence levels are based upon estimates of sources of variance. The sampling protocol is an illustration using the available data. However, without going and collecting some pilot data to estimate the sources of variance for a particular technology, using the estimates obtained from the available datasets gives a good indication of how future sampling plans should be conducted. Other technologies in other climatological regions may yield different results.

An additional NSF test dataset—ETV—was obtained in March 2004 after all of the data had been analyzed. A comparison was made between this new dataset and previous NSF datasets to see if the new data would affect the earlier relationships. The new dataset was only for the Bioclere technology.

Table 2-8 gives the summary statistics for this ETV dataset along with the previous statistics for the NSF test site data and the residence data. Figures 2.27 and 2.28 in Appendix A, show the NSF test dataset and the new ETV dataset. Means and medians are not particularly different between the two datasets. The within variance between the two test center datasets is substantially greater for the ETV dataset. When the underlying test conditions are evaluated, flows and influent concentrations between the two protocols are virtually identical, as are sampling procedures, and others.

The major difference in the protocols is the length of time that the two different tests run: NSF Standard 40 runs for about six months, while ETV runs for 1.5 years. The added time not only allows the treatment technology to experience incoming wastewater characteristics that have the potential to be different, but also the ambient temperatures in the longer test cover all seasons of the year. While temperature is generally acknowledged as having a significant influence on nutrient treatment, potential effects on suspended solids removal and BOD reduction may be significantly less, yet easily observed, if present. These two factors (potential increased range of characteristics of the wastewater and the seasonal influences of temperature) could affect the performance of the treatment units, presumably increasing variability within the performance of an individual unit. Further research is needed to determine if the length of the test center evaluation will have a significant difference in the technologies' performance evaluation.

Table 2-8
Bioclere, BOD, and TSS Summary for Residences, NSF, and ETV Datasets (mg/L)

	Residences	NSF	ETV
BOD			
Mean	23.15	11.1	15.2
Median	12	11	10.1
Within Variance	272.2	20.0	148.0
TSS			
Mean	22.1	5.3	17.1
Median	10	5	11
Within Variance	154.9	5.0	248.1

Summary of Tasks I, II, and III

The purpose of this project was to provide assistance to industry, regulators, and professionals who use science to make decisions regarding approval of advanced technologies. Objectives 1 to 4 as stated in Chapter 1, relate to the statistical analysis that was done in this chapter. This consisted of examining the relative value of field and testing center data, determining the relationship of the different datasets, and attempting to present procedures (such as sampling protocol) for professionals to collect adequate field data capable of verifying the performance of technology.

In summary, the lognormal distribution generally fits the residence datasets better than the normal distribution, due to large values in the datasets. Therefore, applying a log10 transformation to the datasets gives better estimates of the variance. Without applying a transformation to the data, the median may be a better estimate of centrality than the mean, which is affected more by extreme large values.

If the mean and the median are similar in magnitude for a dataset, this indicates that the dataset is symmetric and that a transformation of the data may not be necessary. Statistical analysis showed that there was no obvious inter-annual trend. Consequently, for the technologies evaluated, random future sampling at a residence for BOD and TSS can be started and finished at any time of the year. This may not be so for other temperature-affected parameters (such as nitrogen) or for different technologies.

The statistical analysis concluded that the variability associated with test center data was significantly less than the variability of data collected from real world situations. Therefore, the two data distributions are dissimilar and one data distribution set (such as test or field) cannot be used to accurately predict the other. This conclusion may bring into question the value associated with either data type for future evaluations. In fact, both data types have high value for their respective purposes. An independent, trusted screening process (such as test center) is needed to make sure that any technology installed has a reasonable chance of being successful. In addition, field testing (as described in the following) is needed to help predict what the real-life performance of the system is going to be.

Generally, the largest source of variability is between residence sites rather than within sites. A sampling plan with more residence sites and fewer samples at a site will be more efficient in reducing the variance of the overall mean, since the standard error of the mean is reduced considerably by increasing the number of residence sites and is only scarcely affected by increasing the number of samples per residence site. On the log10 transformed scale, the standard error of the overall mean can be reduced if four samples are taken at a residence site. Taking any more than four does not significantly reduce the standard error. Thus, a sampling protocol in which four samples are taken in a random sampling plan with more residence sites and fewer samples at a site will be more efficient in reducing the variance of the overall mean.

A sampling plan should evaluate as many locations as necessary to achieve the desired standard error of the mean. For the study parameters, a sampling plan for 7 locations will provide for 84% confidence, 18 locations for 95% confidence, and 37 locations for 99% confidence. This suggested sampling protocol can be used by others for studies (provided that the standard error of the mean corresponds to the data in this study).

These conclusions regarding a sampling protocol cannot be generalized to other technologies for other environs or other variables of interest. This study develops a method for establishing a future permitting sampling protocol if the two sources of variability are known for your situation, namely between-residences variability and within-residences variability. Thus, one needs to conduct a pilot study to estimate these sources of variability and apply similar analyses as done for this project.

3 USING THE WEIGHT OF SCIENTIFIC EVIDENCE IN REGULATORY DECISION-MAKING

Introduction

This section presents the development of a Decision Support System (DSS) that can assist with data assessment for determining appropriate regulatory decisions for onsite technologies. The DSS is based on a novel technique that allows the user to "weight" various forms of scientific evidence to assess and compare results from different types of studies of a technology. Science and regulatory decision-making is discussed first by introducing the scientific method and important pieces of a strong scientific foundation for decisions. Then, the concept of how to weight scientific evidence is covered, with special attention given to typical detours to be aware of when making scientifically-based regulatory decisions. Finally, the DSS is introduced and instruction given regarding how to use rating scales (such as, weighting methodology) included in the DSS when assessing and comparing the types and amounts of studies that are needed to determine if the scientific weight of evidence supports a regulatory decision.

One of the benefits of this approach is that the DSS provides regulatory agencies and manufacturers with a method to determine the issues of concern even before assessment of actual datasets for regulatory decision-making. Furthermore the process of using the DSS provides a clearer understanding of the assumptions made regarding the value and weights given to differing types of datasets when making regulatory decisions.

The DSS developed here is broad-based and intended to help improve scientific-based regulatory decision-making in general. Therefore, the description of the DSS and how to use it goes from the general (all types of onsite wastewater regulatory decisions) to the specific (that is, the assessment of pretreatment technologies).

The endpoint goal for the demonstration of the DSS was to determine how pretreatment systems perform in the field in the long run—an important issue for regulators. Therefore, the process used here for the DSS addresses that particular question. Bear in mind, however, that this process will also be pertinent to other questions regarding regulatory decisions in the onsite wastewater field.

The specific project objective regarding pretreatment technologies as discussed in Chapter 2 is used to demonstrate how to use the DSS. The determination of the amount and quality of data needed to make regulatory decisions regarding removal of Biochemical Oxygen Demand (BOD) and Total Suspended Solids (TSS) by pretreatment systems is the critical "decision endpoint" that must be addressed in this project.

The DSS developed here can be used to assist with many regulatory technology decisions in the onsite wastewater field. In fact, the purpose here was to develop the DSS itself—not to actually use it for the assessment of specific datasets—and help render a decision about pretreatment technologies. One of the factors addressed while actually using a tool such as the DSS could be a determination of how to value or rank the use of field and/or more controlled studies (such as test center datasets) compliance samples to decide on advanced pretreatment units that regulators can rely upon to predict performance in the field.

Specifically, this project involved decisions regarding three types or brands of pretreatment technology systems for two parameters (such as BOD and TSS). This focus is pertinent to achieving the objectives of the project. But, as indicated earlier, this project goal is used to illustrate (through the use of this example) how the DSS process can be used for any regulatory question regarding system technology performance.

One of the first parts of using the DSS process is to define a decision endpoint. Nothing, in the way of data assessments, can be accomplished without clearly understanding the decision that needs to be made once the data have finally been evaluated. Then, once a decision endpoint (described later in more detail) has been developed through an evaluation and review process, a determination is made regarding what types of data are valued by regulators and scientists for reaching that decision endpoint. At this point in the DSS process, one goes from the generic to the specific. That is, one goes from ranking the importance of different types of studies (before even evaluating actual datasets that have been collected) to actually evaluating and scoring the data from specific studies that have been completed and submitted for evaluation. In summary, once the decision endpoint is defined clearly, then the reviewers (state regulators and research scientists for this case) can begin the generic process of determining the value of various types of datasets for making that particular decision.

A statistical assessment or comparison of datasets from test centers and field regulatory compliance samples was provided in Chapter 2. It compared, statistically, the merits of two types of datasets for predicting field performance of pretreatment units. The information generated from this statistical assessment can assist regulators and scientists who use the DSS for this type of data and decision-making.

First, though, how to make the best science-based regulatory decisions needs to be reviewed. This discussion is broader than the direct decision about pretreatment technology performance, in that it describes how science works. A deeper understanding of the workings of science can be valuable to many in the regulatory community. This understanding is particularly valuable those who have not had the opportunity to be involved first-hand in extensive research at the principal investigator level, such as scientists at major research universities and governmental or industry research facilities. Of course, there are numerous regulators who have had extensive personal experience and training in the scientific method and its appropriate role in regulatory decision-making. For that audience, this section will serve merely as a review.

Science and Regulatory Decision-Making

Science and regulatory decision-making in the onsite wastewater field was discussed earlier in an article by Hoover and Beardsley (2000) published in *Small Flows Quarterly*. This discussion is repeated here [with permission from the National Environmental Services Center (NESC)] because of its relevance. Comprehending this material is the underpinning for the remainder of this chapter and for getting proper results from the DSS by any user, no matter their technical expertise in onsite wastewater technologies. Hoover and Beardsley (2000) wrote that onsite regulators and review panels across the country are evaluating a growing number of manufacturers' requests for technology approvals. Technical support documentation for product approval submittals from manufacturers range from peer-reviewed papers with attached third-party research reports, to claims of "our system works just like Product X's system," with little supporting third-party research.

As stated earlier, states and provinces are remaking their entire rules into more performance-based approaches. The growing environmental focus in onsite wastewater is causing a shift in emphasis from the traditional disposal aspect to more of the treatment aspect in rule revisions.

This section discusses the role of science in helping make these decisions in the future. How should a regulator decide about the amount and quality of data needed for a specific product approval request? What place should science have in the larger decisions regarding changing your town, county, state, or provincial rules? As a practical matter, what role should other factors besides "pure" scientific studies play in regulatory decision-making (such as basic environmental values like "how clean is clean?" or how to use data that is not "pure")?

These issues will be looked at from the specific to the general—that is, from individual product approval decisions to decisions necessary for modernizing technology review and decision-making processes. The necessary role of science in those decisions will be emphasized. The scientific method will be reviewed briefly, how to build a scientific foundation for making these decisions will be discussed, and the scientific method and its products will be put into the broader context of practical regulatory decision-making.

The Scientific Method

First, here is a quick review of the scientific method for conducting a research study. These parts of the scientific method go to the heart of science and are the pieces that make science work. In general, for the onsite wastewater field of science, they include six pieces—or steps—as follows:

1. Literature review and problem definition

2. Hypothesis

3. Methods

4. Data collection, data analysis, results, and discussion

5. Conclusions and recommendations

6. Publication

The literature review involves a detailed assessment of the published scientific and non-scientific literature. As the literature review proceeds, the researcher begins to define the parameters of the problem.

A clear definition of the problem helps the researcher develop the project objectives, and hence, the hypothesis to be tested. The performance of the systems of interest (the treatments)—for example, the XYZ trench design—are usually compared experimentally to a control—for instance, the conventional trench—or to an industry or environmental standard. However, note that our understanding of the conventional septic system trench is imperfect.

The research methodology will be developed to control all outside influences to the maximum extent possible, reduce bias, and focus upon the specific project hypothesis. Usually statistical levels of significance will also be defined.

Data are collected regarding performance of the treatments and the control. These data are then verified, analyzed, and tabulated, and the results are developed into the form of tables and figures with explanations. For field research, the methods may have to be adapted and adjusted as one goes from plans on paper to reality in the field.

Discussion places the dataset, or results, in a meaningful context relative to the methods and the existing literature. Here, the potentials and limitations for extrapolation of the results to other environmental conditions are exposed.

The research conclusions follow naturally from the discussion of the results. The researcher summarizes the major findings that can be confidently supported by the data collected during the study. Conclusions from any one study are usually specific and not sweeping or broad unless other supporting studies are cited extensively, such as in a review paper regarding a technology.

Publication of the results is just as important as the research itself. Publication usually begins as a research report or thesis, which is then refined or boiled down into its essence in conference proceedings, peer-reviewed journals, or book chapters. Peer-reviewed research that is published in scientific journals carries more weight than that published without review. The least significance is accorded to raw, unpublished data collected by a manufacturer who can benefit from the results.

Building a Strong Foundation

Building sound scientific-based decisions is somewhat like building a castle from blocks. Start with the building blocks of the foundation and go up from there. The castle is built block-by-block. Each research study or test is a block (or maybe two or three blocks if it is well done).

The best building blocks for scientific decision-making are peer-reviewed, unbiased, controlled, replicated studies that encompass a broad range of conditions. Onsite regulations should be based on these types of studies.

Peer Review

Peer review helps ensure that the foundation will hold up under scrutiny and will support the weight of substantial conclusions. It also helps ensure that the "castle design" is sound. In science, peer review normally occurs at both the proposal stage and at the conclusion of the study to assure that the design is appropriate to the hypothesis tested and that the data support the conclusions stated.

Unbiased Research

Unbiased research includes third-party studies where the participants have little to gain regardless of the success or failure of the system under study. This helps assure objectivity. The results and conclusions are the original scholarship of the researcher and are independent of both the funding organization (usually the manufacturer) and the group asking the questions (usually the regulatory authority).

Unbiased, third-party research assures that the building blocks are true-to-form, that they have square edges that are not slanted always in one direction, and they are joined together properly with other blocks (studies) to build a coherent foundation.

Controlled Studies

The best research studies have a strong measure of control on outside variables, or factors that could add "noise" to the dataset. Otherwise, the researcher might not be able to discern the true impacts of the variables being studied. Statistical designs and research methodologies are used to block out extraneous influences so that the research results will focus on the treatments of interest.

Intuitively, the side-by-side study protocol provides a high degree of experimental control. But, when properly done, the side-by-side protocol is expensive, can only be done at one or two sites, and must be started from scratch with newly installed systems. Soil properties vary so substantially, even in one soil at one site, that there may well be more experimental variability within replications of any one treatment, than between the treatments of interest.

Using the castle analogy, the degree of control in a study affects the internal strength of each building block. The better the control of extraneous environmental conditions, the more substantial and stronger the building block.

Replication

Replication is the key to the size of the castle and hence the breadth and weight of the conclusions about a technology or a proposed rule change that can be held up by the foundation. If relationships between the treatments and the control hold up consistently over many replications within an experiment or between different experiments under broader environmental conditions, then the study results are less likely to be due to random chance and the conclusions become stronger.

Broad Ranging Environmental Conditions

The intensive, highly controlled type of side-by-side study is invaluable for providing clear results and conclusions about system performance at one site. However, the results cannot be easily extrapolated outside the narrow set of conditions tested during the study without additional data. Also, since the side-by-side research process takes time, it is not always the most efficient approach for assessing performance. Waiting until new side-by-side systems can be installed, biomats are enabled to form, and the systems are allowed to mature before each new question about system performance can be answered is not realistic.

But another approach—the field performance survey—can provide research data that complements data from the side-by-side protocol. Statistically-sound field performance surveys that evaluate random, stratified groups of hundreds of real systems under a broader range of environmental conditions, can help the researcher extrapolate the results from the narrow set of conditions in an intensive side-by-side study. Field performance surveys can address a broader range of conditions—soils, climates, ages, wastewater strengths and flows, and system design, installation, and operation variations—that represent reality in the field.

The Scientific Method and Regulatory Decision-Making

The scientific method and the need to build a much stronger scientific foundation as the basis for improving the quality and credibility of decisions in the onsite field have been discussed. But scientific method and studies are not enough in the world of practical regulatory decision-making. Three other considerations must complement the scientific method. To put it differently, consider the broader context in which onsite science must work.

Environmental Values, Not Just Science, Determine Performance Standards

The same objectives need to be considered for a more performance- and treatment-based approach to regulation of onsite wastewater as in all other aspects of environmental protection. How clean are clean groundwater and surface water? Federal law and regulation, of course, have provided one set of answers, such as "fishable/swimmable" surface waters. States and localities will have to decide whether the federally-defined objectives—or values—are sufficient. After those value decisions have been made, good science is necessary to determine whether products, technologies, management practices, or other pollution mitigation mechanisms are practical and sufficient to achieve the standards.

The American Public Does Not Make Decisions on the Basis of Science Alone

Most agree that treatment and performance standards are needed, getting better at using scientific data in making decisions is necessary, and more enforceable procedures for management and maintenance of onsite systems are needed.

Implementation Will Require Increased Public Expenditures

How, exactly, do you convince Aunt Millie that she needs to spend money on a better system in order to protect the environment and more money on maintenance to ensure her system does not fail?

The beginning of the answer is that the environmental niche of onsite wastewater treatment needs to see itself as being similar to all other environmental issues. Pollution is pollution. Allies are needed to address problems systematically. Non-governmental environmental groups are needed—with their amazing historical ability to create public interest and support for addressing environmental problems—to share the same values and commitment to improved onsite wastewater management.

There is never enough “pure” science to make perfect decisions. Since soon after the first Earth Day, every environmental regulator has wished to have third-party, peer-reviewed, replicable, and published studies of every aspect of the pollution problem under consideration. Environmental protection, however, will not be quickly advanced—in the onsite field or any other—if the regulator waits on scientific certainty. Therefore, the perfect cannot be the enemy of the good.

In most parts of environmental protection, a kind of informally accepted set of rules has evolved in response to this situation. The rules are collectively called “acceptance of the scientific weight of evidence.” Regulators should not and cannot afford to throw out data. Rather, they should put data into a hierarchy like the castle example. They can give, for instance, one weight to data supplied by a self-interested manufacturer and another weight to informal surveys by other regulators. Regulators can place emphasis on studies performed by non-academic third parties, and on tests performed in somewhat different countries or climates, and laboratory studies versus field studies versus epidemiological studies, and so forth.

In the onsite field researchers must set higher standards for the scientific documentation expected for product approval and other decisions. But, they should also use all the data at their disposal (imperfect as it may be) for making decisions, and accept that judgments will have to be made about the relative value or weight of these data.

As discussed earlier, the scientific process resembles building a castle out of blocks. Each block is a study of some performance claim in which the researcher is interested. What matters is whether there are enough blocks, of sufficient strength, to verify that a castle can stand. For the strength of the castle, a judgment will have to be made. As with practicing medicine, there will never be absolute certainty, but that should not keep people from making decisions or taking risks on new technologies. As in medicine, old remedies are often not the best remedies.

Despite the complications, science surely needs to be the decision-making tool to rely on in the future, just as it has been in other fields of environmental management. Castles built without science are built out of sand.

The Weight of Scientific Evidence

As explained earlier, Hoover and Beardsley (2000) made the case that regulatory and related decisions in this field should be based as much as possible on scientific information. In this section, a discussion is provided (with permission from *Small Flows* editor Tim Surher) that was first presented by Hoover and Beardsley (2001) in *Small Flows Quarterly,* which builds on the earlier foundation. They argued that while regulatory decisions need to be based upon scientific information, regulators could never expect to have scientific certainty in environmental or health decision-making. Hence, for any given regulatory decision, judgments must be made about the amount of data needed and the level of quality of that data.

Previously, little guidance regarding research data evaluation was available to regulators, product approval panels, or advisory committees (operating under various names such as Technical Review Committee and Technical Advisory Committee). How should scientific principles fit into the decision-making process? In this section, this discussion is extended to encourage further agreement on what constitutes a proper basis for these decisions.

Good science depends on a foundation of data and research, but data are often of varying degrees of quantity and quality—spontaneous vendor assertions about performance, for instance, may have some value but they must be considered differently than the results of third-party, peer-reviewed research. In other words, data need to be weighted, and regulatory decisions are usually based on the relative weight of scientific evidence.

This discussion continues along two paths: First, what are the most common distractions or detours from science-based decision processes? Second, a common sense guide to weighting data is proposed. The onsite community of researchers is challenged to develop a structure for comparing the differing kinds of research data needed to answer particular questions about onsite performance. How best, for instance, can one answer a question about long-term treatment performance of a given technology?

Some Common Detours From Good Science

Making Decisions Based on Weak or Wrong Data

Everyone knows this trap: A vendor submits unverified assertions about how a product works, then demands approval of his or her technology. This data may come from an in-house engineer or designer. Such data typically include little substantiated information about real-world performance, and often simply ask the regulator to agree intuitively with the logic of the vendor's design. Intuition, of course, is not science.

As a former member of the North Carolina Experimental and Innovative System Advisory Committee, Hoover has had the opportunity to see firsthand a range of submittals for new system approvals, including both scientific and non-scientific data. The committee has been exposed to everything from large three-ring binders filled with third-party, detailed research reports, to other submittals with little data, but claims of, "Our system works just like Product X's system." Many onsite regulators, including those involved in this project, have also seen these issues firsthand.

The ideal submittal would, of course, include a range of scientifically reviewed papers and detailed research reports. Far more frequently, the committee is only presented raw data (tables and figures). These data, alone, have limited utility. At a minimum, the data should be summarized and analyzed within the context of the design flows and waste strengths tested, the methods used to collect and analyze the samples, and the relevance of similar literature to the submittal request.

Earlier, the scientific method and the six pieces of the scientific method that make science work were discussed. These are:

1. Literature review and problem definition
2. Hypothesis
3. Methods
4. Data collection, data analysis, results, and discussion
5. Conclusions and recommendations
6. Publication

Product approval submittals that include only in-house tabular lists of performance data—without any of the other five pieces of the scientific method discussed above—are only one-sixth of what is ideal and hardly enough for a good decision. The only reasonable option left to a technical review committee is to ask the manufacturer for more complete information or studies.

Even worse, however, are submittals based purely upon statements that a technology is "equivalent" to another without any substantiated supporting research other than some design drawings. These approval requests are simply hypothetical assertions; they have little relation to professional presentations of scientific data upon which reasonable decisions can be made.

Throwing Away Data Because They Are Not "Perfect"

If, from now on, decisions are made only on the basis of peer-reviewed academic studies, many decisions are unlikely to be made. Vendor submissions of long-term field performance data—especially if that information is not contradicted by extensive reported failures—have use even if not peer-reviewed. Similarly, extensive field data and experience may be useful in adding weight to a decision even if that data is not gathered through rigorous application of the scientific method in a controlled study.

Technical review committees have also seen good data thrown away while the ideal study is pursued. As an example of this from one state, a long-term, multi-year study of a pretreatment technology that was installed at a wastewater treatment plant was considered invalid by a regulatory agency. Although the samples were analyzed by an independent third-party laboratory, a third party did not physically carry the samples to the third-party testing laboratory (within the same facility). In essence, this dataset was declared invalid because it was not perfect and assigned zero value relative to answering the technology performance and regulatory question at-hand. Certainly the issue of sample handling chain-of-custody was potentially important and relative to the data reliability, but should an entire dataset of a multi-year study be just thrown away without further thought and consideration of alternatives?

After further reflection, the regulatory agency involved chose to make use of the study. However, it required that additional samples be collected for two more months by a third party to confirm that the initial results were reliable and consistent with results from the later-collected samples. Science does not give perfect answers, particularly using one study, and scientific studies themselves are rarely perfect. This is particularly true concerning highly variable natural and biological environments. More often than not, nature acts in terms of accuracy as opposed to precision.

All non-fraudulent data can be useful; the challenge is deciding how useful, and how much data of what quality are necessary to make a responsible decision. Determining what data are fraudulent can be a difficult task, but the DSS proposes that all non-fraudulent data have some merit, and that merit will be assigned by the combined wisdom of the expert panel. As in the preceding example, imperfect data can be useful for decision-making, especially if supplemental or confirmatory documentation is provided.

Several suggested approaches to determine fraudulent or non-fraudulent data are offered as follows:

- The reviewing expert panel or state regulatory agency could independently spot sample or split samples by a separate independent entity during the study's development, its implementation, or as a follow-up to the study. This could also be built into the scope of the original study and its Q/A procedures.
- Laboratory testing accuracy and veracity could be checked by a split-sample lab test either as a percentage of samples taken or by requiring a certain number of split samples between two different labs. Laboratory testing has not been reliable as evidenced by several state certified labs in Pennsylvania being decertified due to fraudulent reporting of lab results.
- The expert panel should question the reliability of data, its collection, and laboratory analysis in respect to each dataset and study. Questions such as: why are there so many samples with the same precise reported value, data reported at or below detection method levels, or where there are wide variations in comparison with data from other studies and data sources.

Reporting, keeping all data, and explaining its variances remain the standards for good scientific methods. After all the statistics, the data relationships, and trends developed will point toward any less-than-accurate and reliable datasets.

The Subtle Detour: Substituting Unsupported Assumptions for Good Science

Decision-making, like nature, abhors a vacuum—when the science is inconclusive, human tendency is to fill the gap with human beliefs. Consider the following two examples:

- Pretreatment systems are widely assumed to be the next logical step in furthering onsite system performance. This may be so, but based on what conclusive set of studies? Are these technologies more effective in dealing with viruses, nitrates, and other contaminants? What happens to the soil treatment process below the drainfield if the trenches no longer have a biomat because pretreatment reduces the waste strength?
- Sidewall is (or is not) a crucial component of system performance. Based purely on the wide variety of positions on this issue, as expressed in state regulations, it seems that pretty clear guesses are being made rather than science-based decisions on the value of sidewall. There has been some excellent research by past researchers on sidewall effects; but has it been forgotten?

Committing to the role of science means watching out for the inclination to include unsupported assumptions, even if those assumptions are made in the noble belief that environmental protection is being improved. It may not be.

A Proposed Simple Guide for Ranking (or Weighting) Evidence

Hoover and Beardsley (2001) suggested a simple guide for weighting datasets of different types. At one extreme is the inventor who says his new product will address all treatment and disposal needs at a cost even your children can afford. At the other end is the perfect laboratory study, peer-reviewed, with results published in juried journals. Consider this rather simple structure, or something like it, suggested by Hoover and Beardsley (2001) for using and weighting available evidence.

1. Unsupported performance assertions by a vendor = 0 points

2. Vendor submissions of field performance over time in one state = 1 point

3. Vendor submissions of field performance over time in one region = 2 points

4. Vendor submissions of field performance over time in many states = 3 points

5. Third-party studies (such as by regulators themselves) of field performance over time = 4 points

6. One peer-reviewed, published, third-party study of a performance claim = 5 points

7. A confirmatory study, with the same or similar results = 6 points

Because the data submitted are usually less than ideal, some in the regulatory community are beginning to search for the perfect research protocol or study design that will answer all questions, or at least answer enough of the questions to facilitate sound decision-making. But, realistically, can one study be designed that would answer all the questions? Is there a perfect study protocol that will, in one fell swoop, provide enough data to make sound decisions? Scientific research is a continuing, iterative process wherein a little is learned in each step. Always be flexible and willing to adopt new information as it is generated and confirmed.

Consider the analogy from Hoover and Beardsley (2000) regarding the building of a castle. Scientific consensus never occurs based on one perfect study—at the minimum, that study would have to be confirmed. Rather, the scientific castle is constructed brick-by-brick (study-by-study) until the weight of evidence is strong enough to hold up the claims asserted about that technology's performance.

Hoover and Beardsley (2001) suggested that the onsite community should develop a comparative assessment of the qualities and strengths of different types of research studies. These studies could range from detailed laboratory studies using one soil and one quality of effluent to broader, in-field survey assessments that investigate the performance of hundreds, or even thousands, of systems. It is only through comparisons of different types of studies that an answer can be found to the question of what kind of research (as well as how much scientific data and of what quality) is needed in order to make scientific decisions about technology approvals. The following DSS method is an outgrowth of the recommendations for development of tools and methods to enhance the rather simple weighting structure first proposed by Hoover and Beardsley in 2001 as a starting point for assessing datasets in a more scientific and defensible manner.

The following method describes a Decision Support System (DSS) for assessing datasets or research studies. It introduces a process for making science-based decisions. Some of the first steps include comparing research studies and different types of datasets and ranking how valuable they are for a particular regulatory decision (that is, a specific decision endpoint). The weights for eight different data attributes are established. These assessments are made prior to actually looking at the data from a particular study. Following these assessments, then the research studies are evaluated and scores given to the data.

Decision Support System (DSS)

The primary references used in developing the Decision Support System (DSS) and in using the weight of scientific evidence approach were Mass DEP (1995), Hoover and Beardsley (2000), and Hoover and Beardsley (2001). Note that while the Mass DEP (1995) publication was used for developing a strategy for this DSS, the approach used here is substantially different than that used by Mass DEP (1995).

The Mass DEP (1995) approach is one in which the user quantitatively or qualitatively assesses the weight of scientific evidence needed for regulatory decision-making. While the approach used for the DSS is similar in some respects to the Mass DEP (1995) method, and acknowledgement of their method as a resource is given here, the approach used for the DSS takes a different direction. It does, however, borrow substantially from many of Mass DEP's concepts, sometimes using the concepts in a different way. In essence, the DSS blends the Mass DEP (1995) approach together with that suggested by Hoover and Beardsley (2001).

Overall Approach

The overall approach used in the DSS developed herein consists of seven steps that need to be followed sequentially:

1. Setting a final score value suggested as adequate for regulatory decision-making

2. Determining and defining the decision endpoint

3. Ranking, numerically, each of 10 different types of studies (for example, datasets) that could be used to address that particular endpoint before assessing data from any specific research study

4. Assigning numerical weights for each of 8 data quality/quantity attributes relative to that particular decision endpoint prior to assessing data from any specific study

5. Evaluating data from specific studies to determine the value of that data for each of the eight data attributes for the decision endpoint and giving the data numerical scores

6. Summing the results of the calculations for data value and weight for each study conducted and submitted for assessment of that technology

7. Comparing the calculated scores to the predetermined set of standard values suggested in Step 1 for regulatory decision-making

Each of these steps will be described later. However, first some concepts will be introduced about dataset properties and the value of different types of data before describing how to use the DSS process in detail.

Dataset Properties and Data Value Characteristics

First, it is important to remember, as discussed earlier, that all data are valuable as long as they are not fraudulent (Hoover and Beardsley, 2000 and 2001). However, all datasets do not have equal value for decision-making. This could be the case because the type of study is not appropriate for making a particular regulatory decision, or there may be data quantity or data quality issues in the actual research study dataset.

Even for comprehensive, highly-controlled data about a pretreatment technology (such as data submitted from test center studies), questions remain about how well the test center data predict long-term performance in the field. Hence, what weight should be given to that type of data?

The previous statistical assessment in Chapter 2 provides substantial insight into the relative value of highly controlled, but limited, test center data versus less-controlled, but field performance data from a larger number of sites. The first two specific questions, therefore, are:

- What relative ranking value should be assigned to particular types of datasets (that is, the types of studies conducted such as the test center datasets versus field performance datasets evaluated in Chapter 2 earlier)?
- How much weight should be given to different data quality and data quantity attributes for enhancing dependable decision-making?

In the onsite wastewater field there are many different types of datasets. Any of these could be of some value for the decision endpoint for this project, which is to "determine how pretreatment systems perform in the field in the long run." The types of datasets presented to regulators for this decision regarding a particular pretreatment technology could potentially include any of the following datasets (Table 3-1).

These are not presented in any order of value or weight. But some, if not many, of these types of data have been submitted or used for decision-making. Any of these study types are likely to be submitted for proposed regulatory decisions for any technology-based regulatory approval request (nutrient reduction technology requests, substitute trench media drainfield area reduction requests, dispersal system technology requests, tank additive approval requests, and others) well beyond the fairly limited question of pretreatment technology BOD and TSS performance.

Each of these datasets has properties that influence its reliability or perceived reliability during a regulatory decision-making process For instance, a field-based study that has been published in a refereed journal (Type A) will be valuable for making certain regulatory decisions. The quantity and quality of the data in such a study will be substantial, and there will be excellent experimental control. Compare this to a university research report that has not been peer-reviewed, such as:

- An ASAE paper or a final project report (Type C)
- An exploratory dataset developed by a university researcher (Type D)
- A test center dataset such as an NSF-like or ETV-like study (Type E)
- A state regulatory agency dataset consisting of numerous, but relatively uncontrolled, sampling results from pretreatment system compliance sampling or system hydraulic performance compliance observations by operators (Type G)

Table 3-1
Ten Types of Studies (Datasets) Submitted for Regulatory Decision-Making

A.	Field datasets that are published in refereed journal articles that have a high degree of scientific rigor and use generally accepted scientific methods.
B.	Laboratory and bench-top datasets that are published in refereed journal articles that have a high degree of scientific rigor and use generally accepted scientific methods. These are the same as Type A above except they are conducted under artificial conditions.
C.	Independently published university and governmental research reports that are not peer-reviewed, but that generally have a high degree of scientific rigor. This might include ASAE symposium papers, final project research reports for a state funding agency, and more.
D.	Exploratory datasets developed by independent university or government researchers. This might be a non-published dataset of the performance of three or four systems that are assessed using minimal resources in an unfunded study.
E.	Test center datasets developed using specified protocols, such as NSF, ETV, and Environmental Technology Initiative (ETI) (might also include DelVal datasets, Florida Keys project results and Michigan State University test center datasets), that are highly controlled and use “homogeneous” wastewater of a specified quality, usually from a treatment plant.
F.	Demonstration project datasets such as NODP, National Community Decentralized Wastewater Demonstration Project, or EPA/319h projects that have varying scientific rigor, often not as substantial as test center datasets, but usually more field-based using sewage from real homes.
G.	State and county regulatory datasets that may be primarily composed of compliance data samples. This might include results from pretreatment technology samples collected yearly from hundreds or thousands of real-life systems. Datasets taken by Responsible Management Entities (RMEs) or maintenance firms can also be considered.
H.	Vendor-developed journal articles that are peer-reviewed.
I.	Vendor-developed research reports that are not peer-reviewed.
J.	Vendor-developed datasets, including exploratory datasets that are developed by product manufacturers, inventors, and the like. This might include extensive lists such as an Excel spreadsheet of pretreatment sample results with little or no critical analysis of the results.

The Type A study might be ranked as more valuable than the other types of datasets for making a decision about the performance of a dispersal technology, such as LPP or drip dispersal. The Type E dataset might be viewed as comparable to Type C, but more valuable than Type D and G datasets, for that type of decision. On the other hand, when evaluating the performance of a pretreatment technology, say for BOD reduction, there may be a different assessment of the value of these datasets. In such a situation, the Type G study (compliance sample dataset) might be, if it contains results from hundreds or thousands of real-life systems, ranked as having equal value to a Type A study (peer-reviewed journal article) and certainly have as much or more value than a Type E study (NSF-like assessment of one unit).

In addition to quantitatively ranking the value of different types of datasets for the regulatory endpoint that is being evaluated, the DSS will help to numerically weight each of the data quality/quantity properties.

Table 3-2
Eight Dataset Properties That Can Influence Reliability of Decision-Making

Dataset Property	Examples of Data
Performance Data	• Influent BOD – Mean, median, range, s.d., CV, *n* • Effluent BOD – Mean, median, range, s.d., CV, *n* • Model #, design type of unit, design progression stage • Sample collection – Grab versus composite o Time composite versus volume composite o Time of grab sample – Chain of command of sample
Flow Data	• Daily flow (and how measured, such as meter reading, # of people) – Mean, median, range, s.d., CV, *n* • Monthly flow – Mean, median, range, s.d., CV, *n* • Peak flow limits – Upper range, period • Number of people or bedrooms • Design flow
Replication	• Number of sites/systems assessed in one study • Number of samples collected from each site/system • Replication of lab analysis results (QA/QC) • Number of separate replicate studies • Confirmatory study

Table 3-2
Eight Dataset Properties That Can Influence Reliability of Decision-Making (Cont.)

Dataset Property	Examples of Data
Experimental Control	• Experimental control or standard used for comparison • Protocol, written, peer-reviewed • Degree of control over outside environmental conditions that could influence performance • Statistical assessment • Certified laboratory/ standard recognized method (SOP) • Quality Assurance Project Plan (QAPP) written and followed – Gross “failure” causing system to be withdrawn from study
Range of Environmental Conditions Tested	• Climatic conditions – Monthly avg. temp during testing – Yearly avg. temp – Precipitation during testing, daily, monthly, yearly as appropriate – Comparison of test precipitation conditions to long-term averages – Mean annual soil temperature (MAST) • Soil conditions • Map unit or soil series • STATSGO data for that series/map unit • Onsite soil morphology data • Onsite K_{sat} and other measurements • Landscape position • Range of environmental conditions tested during study – Flow, wastewater strength, siting variations, design variations, installation variations, operational variety, ability to extrapolate dataset to a broad range of environmental conditions • Distribution of environmental characteristics
O&M Conducted During Study	• Did they have an O&M schedule?

Table 3-2
Eight Dataset Properties That Can Influence Reliability of Decision-Making (Cont.)

Dataset Property	Examples of Data
Third-Party Assessment	• Vendor data • Demonstration project dataset • Third-party test center data • Third-party researcher • Degree of independence • Government collected data
Peer-Reviewed Data/Study/ Publication	• Peer-reviewed research protocol • Peer-reviewed study • Published by other than vendor • Peer-reviewed publication (journal)

Table 3-2 includes some of the key data properties evaluated in the DSS process. The DSS helps a regulatory agency determine which of the properties in Table 3-1 and Table 3-2 are most valuable for the decision at hand and therefore should receive the greatest weight. In addition, it will become clear that by making this a quantitative open process, this DSS method, by default, clarifies the assumptions made by scientists and regulators about the value and weights assumed for each of the data attributes.

Note that while ten study types have been identified, there are somewhat arbitrary distinctions between these. Arguably, some could be grouped together to have fewer types of research studies or datasets. Some could be further subdivided or split to create an even larger number. These ten dataset types, however, reflect the best effort at defining different types of datasets that would likely have distinctly different numerical rankings for scientific decision-making in the onsite wastewater field. Likewise, eight different attributes of data have been identified that can be weighted relative to their importance in decision-making. These could also be grouped or split depending upon preferences.

Having introduced the ten study (or dataset) types to numerically rank and the eight most critical data attributes to numerically weight, now the Decision Support System (DSS) and how to use it is described.

Decision Support System Introduction

It is important to understand that when using the DSS, decisions are made *by* regulatory agencies, not *for* them. The DSS is a decision *support* system, not a decision-*making* system. There is no magic here; the regulatory agencies themselves (working with a panel of experts that includes, at least, some onsite wastewater scientists) determine the value of data characteristics. But the DSS process aids in that determination.

When the DSS is used as designed, it can help to assure that there are no unsaid assumptions about the amount of weight assigned to different data quality/quantity attributes. When actual study results are assessed, the DSS helps regulators determine quantitative scores for the data by breaking the scoring of the data down into the eight data attributes. Throughout this entire process, the DSS uses a quantitative approach giving quantitative rankings based upon the type of study, as well as quantitative weights for data attributes and numerical data scores for the actual research data.

The DSS assists users in evaluating appropriate rankings for differing types of studies (such as datasets). It does not provide the data rankings, but helps regulators and scientists develop their own quantitative rankings for different datasets based upon their understanding of the research process and perceptions (biases) regarding which attributes of the data are most valuable for regulatory decision-making.

An expert panel approach is used wherein a series of questionnaires in the form of Excel spreadsheets are given to the interested parties (onsite technology regulators and onsite wastewater scientists in this case). The DSS approach provides a self-assessment weighting tool to facilitate determining the value of different types of data quality and quantity attributes.

As a DSS is used in real life, a group such as the NEIWPCC Project Team can serve as the core of the expert panel that will develop rankings for different dataset attributes. However, it is recommended that other regulators from the State Onsite Regulators Alliance (SORA) and/or regulators and scientists from the National Onsite Wastewater Recycling Association (NOWRA) are also involved to a large extent. Ultimately, the use of this DSS could ideally be institutionalized within a national group such as SORA, NOWRA, NEIWPCC, National Environmental Health Association (NEHA), NSF, or NESC. This is the best approach, since the entire DSS process is fairly complex and may require more effort than some state regulatory agencies can handle. It would also be advantageous to get a national perspective on the decision-making rather than have separate expert panel evaluations in each state.

Having said that, a national approach would provide an ideal situation from many perspectives. An individual state or local regulatory jurisdiction could quite easily use the DSS on their own after being trained in its use with the input, guidance, and assistance of a local university scientist. The larger the expert panel, the more dependable the results may become. But that will not always be the case. There is no outright requirement for having 8 or 10 expert researchers or onsite wastewater regulators before the DSS can be used. Many local regulatory agencies may not have the staff resources or time to easily conduct the scientific assessment required using the DSS. However, just using the DSS itself to the greatest extent possible locally should help, even for a state where the regulatory agency staff resources are stretched thin. One good local scientist could work with the regulatory staff to guide a state-level (or local) technical review committee through the process and utilize the DSS spreadsheets to make the calculations needed for a technology assessment.

The complexity that is present within the DSS is a result of the inherent complexity of science itself. This is critical to realize. The process of using more scientific approaches to regulatory decision-making injects added complexity into rule-making and technology approvals. But the DSS should bring more objectivity, dependability, and defensibility for the outcomes. Therefore, local and state regulatory jurisdictions are encouraged to look upon the DSS as a method to enhance what they are asked to do by society and to use it to inject a more scientifically rigorous approach into their regulatory decision-making process. The end result should be better decisions for the clients they serve and clearer guidance to manufacturers and others about expectations regarding submittals for product approval requests.

At the same time that the DSS is being used locally, a national group, such as NOWRA, could potentially begin the process of maintaining a permanent database of study results and DSS assessments. That type of database could lead to a long-term cumulative DSS that could be maintained nationally but “tapped into” quite easily by local regulators.

If that approach is used nationally, then as additional studies are completed or datasets regarding a particular technology become available anywhere in the nation, these datasets could be submitted to a national organization and quantitatively assessed using the DSS. The scores determined from those assessments could then be added to the on-going running total score for that decision endpoint to help make better science-based decisions regarding potential upgrades to approvals (for example, from “piloting” status to “innovative use” status to “accepted” status).

The expert panel approach used in the DSS process here was adapted from the Mass DEP (1995) method. The expert panel is used to assess, rank, and assign numerical weightings for each of the eight data properties, as was suggested earlier by Hoover and Beardsley (2001). These numerical weightings vary depending upon the type of “decision endpoint” that is being evaluated.

Use of an expert panel assumes that each regulator and scientist on the panel independently assigns study ranks, data attribute weights, and data scores without undue influence from others. Obviously, after the first round of assessments, discussion will ensue within the expert panel and result in adjustments to the numerical values used. But the process will expose assumptions regarding the values of studies and data attributes.

The numerical values from these assessments are then summed and averaged to determine the value of different data types and quantitative scores for that particular decision endpoint.

The goal is to eventually have a decision-making group that is as geographically large as possible to facilitate consistent decision-making across jurisdiction boundaries. For that reason, a national or regional group such as SORA, NOWRA, NEIWPCC, NSF, NEHA or NSFC are suggested to lead the effort.

The rankings assigned to the ten study types (for example, datasets that could possibly be submitted for decision-making) and the numerical weights specifying the value for each of the eight data attributes are assigned in advance. This allows a determination on approximately the amount and quality of data is needed for a decision prior to the actual assessment of the research data itself from individual studies.

The regulatory agency is then able to specify in advance some standards or expectations regarding the data quality and quantity needed for an anticipated decision. Then, the DSS method helps regulators determine whether the actual data submitted supports the proposed decision.

This approach can help the regulatory agency in a number of ways. It helps regulators clarify and quantitatively describe their decision-making process. It allows them to predetermine the value for different datasets (study types) and data quality/quantity attributes.

Using these values, they can provide guidance regarding how much of what type of data needs to be submitted before a decision. This can and should be done before the actual data is assessed for a specific decision endpoint. Seven steps were identified earlier for the DSS process. Each will be described and illustrated in detail following an introduction to the organization and content of the DSS spreadsheets.

Organization of the DSS Spreadsheets

The DSS itself consists of a series of 15 spreadsheets organized into three categories as indicated in Table 3-3.

Table 3-3
Summary of Spreadsheets Included in the DSS

Sheet	Spreadsheet Page	Description:
		(Sheets 1-3 are filled out by each expert panel member. These are the only sheets filled out by the panel members. Multiple copies will be needed for each panel member)
1	Input Study rankings	Ranking levels for 10 study types by each expert panel member
2	Input Attribute Weights	Data attribute weights assigned to 8 data quality and quantity attributes for each study type by each expert panel member
3	Input Data Scores	Data scores assigned for each of 8 data characteristics based upon a specific dataset reviewed by each expert panel member
4	Performance Weight Summary	Summary of performance data attribute weights assigned by all of the expert panel members
5	Flow Data Weight Summary	Summary of flow data attribute weights assigned by all of the expert panel members
6	Replication Weight Summary	Summary of relative weighting assigned for replication in a study assigned by all of the expert panel members
7	Experimental Control Weight Summary	Summary of relative weighting assigned for the importance of experimental control in a study assigned by all of the expert panel members

Table 3-3
Summary of Spreadsheets Included in the DSS (Cont.)

Sheet	Spreadsheet Page	Description:
8	Environmental Conditions Weight Summary	Summary of relative weighting assigned for testing a broad range of environmental conditions in a study assigned by all of the expert panel members
9	O&M Weight Summary	Summary of relative weighting assigned for the importance of having the same O&M conditions during the research as in the real life use of a system as assigned by all of the expert panel members
10	Third-party Weight Summary	Summary of relative weighting assigned for the importance of third-party data collection and analysis in a study assigned by all of the expert panel members
11	Peer Review Weight Summary	Summary of relative weighting assigned for the importance of peer review in a study assigned by all of the expert panel members
12	Ranking Compilation	Ranking compilation of study types (type of dataset) by expert panel (for example, taken from sheet 1 for each panel member)
13	Data Weight Compilation	Summary of calculated relative data weighting assigned by the expert panel for all study types (for example, summary weights taken from sheets 4-11)
14	Data Score Compilation	Summary of data scores assigned by the expert panel for a particular study for all studies submitted or measurement endpoints assessed (for example, taken from sheet 3 for each panel member)
15	DSS Calculation	Calculation of expert panel summary scores for all studies submitted or measurement endpoints (“ranks”, “weights” and “scores” taken from sheets 12, 13, and 14, respectively, and final calculations made to determine total score for the decision endpoint)

The DSS is presented in a Microsoft Excel format and is provided on the CD Resource Tool. The DSS is accompanied by this text, sample spreadsheets, and a Microsoft PowerPoint overview on the use of the DSS.

The first series of spreadsheets (1-3) are used by individual expert panel members for recording their study ranks, data attribute weights, and data scores. The second series of spreadsheets (4-11) take the data weighting values from individual panel members and use them to calculate weights for the entire panel for each of the eight data quality/quantity attributes. The final series of spreadsheets (12-15) include the summary sheets and calculations of scores for measurement endpoints and the ultimate decision endpoint. These include #12, which is a summary of the expert panel rankings for each study type; #13, which is a summary of the calculated weights assigned by the panel; #14, which is a summary of the data scores assigned by the expert panel; and #15, which provides the details for the calculations of the measurement endpoint scores for each study and then adds these scores to determine the final decision endpoint score for decision-making.

Here are the details of the seven steps that must be followed when using the DSS.

Step 1: Setting a Numerical Score for Decision-Making

Since this is a quantitative process, the ultimate decision depends upon setting a final score needed to support a determination that the scientific weight of evidence has been achieved. In many regulatory jurisdictions the approval process is basically a yes or no decision. That is, once the technology is first approved, hundreds and thousands of the systems can then be installed without any other assessment.

But in a substantial number of jurisdictions, the approval process uses a graded approach wherein approvals are issued in increasingly substantive levels of acceptance, or stages of approval. As a result, it is necessary to identify different types of regulatory decisions. Of course, each regulatory jurisdiction has its own process and stages of approval. These are simplified and summarized here into three basic types of approvals or regulatory decisions as follows:

1. Approved for controlled limited piloting (piloting), which would allow restricted installation of “less than ten to tens” of test systems depending upon the basic category of technology type.

2. Approved for more extended use (extended use), which would allow “hundreds to thousands” of innovative systems depending upon the technology type, but that also includes numerous stipulations in the approvals regarding use of the technology beyond that required for typically approved systems.

3. Approved as an accepted technology (accepted), which allows general usual and customary use of the technology throughout the regulatory jurisdiction without special approval stipulations (other than those needed to assure continued operation and maintenance).

Each of these approval levels goes beyond the basic experimental approval that allows testing of technologies at one or two sites. As one goes down the approval list, each stage of approval requires increasingly higher levels of confidence regarding technology function and performance. The level of data needed to support each of these stages of regulatory approval should be determined in advance. Suggested appropriate scores needed for each of these approval levels are as follows:

- Piloting/Testing Confirmation Use – 6 points
- Extended Innovative Product Use – 12 points
- General Accepted Use – 24 points

The scores required for the different levels of approval are based upon the expected point levels from various studies. The method to determine point levels is explained later and included in the DSS spreadsheets. However, all of the approval levels above assume that a perfect study that shows excellent system performance is worth a maximum of 8 points.

The transition from piloting to extended use is substantial (a doubling of needed final points), but the level of science that properly allows accepted use of a technology is double that again. Hence, for a technology to be accepted, the approach used here requires three “perfect” studies of the correct type, each confirming that the technology works properly. Alternatively, approximately six substantial studies that are of less than perfect design could suffice, but each must have positive results and address the most pertinent attributes of data quality/quantity.

Step 2: Defining a Decision Endpoint

As indicated earlier, an important part of this process is to determine carefully the questions that must be answered about a technology so that a decision can be made. This is the “decision endpoint.” Each decision endpoint must be thought out and described carefully as it will influence the rest of the DSS process when evaluating a technology. This was noted earlier and it was stated that the decision endpoint for this DSS demonstration was to “determine how pretreatment systems perform in the field in the long run.”

Table 3-4 includes a few examples of other hypothetical types of “decision endpoints” (the technologies are hypothetical). These examples are broader than the specific question regarding pretreatment technology unit performance being considered here, but are necessary to illustrate one of the key milepost steps in using the DSS. It is important to appreciate how much the decision endpoint could change the assessment of even one specific dataset. Hence, carefully determining the decision endpoint is a key milestone in the process.

Table 3-4
Hypothetical Decision Endpoints Used to Illustrate How the Decision to be Made Will Influence the Value of Different Data Quality Attributes

The “Biohiggins” pretreatment technology	Will provide BOD_5 and TSS effluent levels 90% of the time (less than 30 mg/L when used at a single-family home)
The “Coureventor” trench media	Will allow a 30% reduction in drainfield trench length when used with septic tank effluent at a single-family home
The “Bowerseptic” unit	Will reduce nitrogen loading to the drainfield to less than 10 mg/L nitrogen when supplied with any septic tank effluent from a home or commercial facility
The “Grovemaster” pretreatment unit	Will allow reduction in drainfield length by 50% when used on all soils
The “Hepnamiter” septic tank additive	Reduces soil clogging and allows use of previously unusable clayey soils for septic systems

The decision endpoint selected will influence the relative weights assigned to the eight data attributes and the applicability (or numerical ranking) of different types of datasets or research studies. Being specific about how to develop a precisely defined decision endpoint and looking back at the preliminary decision endpoint for this DSS demonstration (determine how pretreatment systems perform in the field in the long run), it becomes clear that more specifics are needed in order to make decisions. It is assumed that there is currently some data available regarding this technology and that it will be assessed whether that data is adequate for dependable decision-making or how much additional data (and of what quality) must be collected for such a decision.

For this DSS demonstration, the plan is to determine the conditions of interest as a reduction of BOD_5 and TSS to less than 30 mg/L by the pretreatment unit. It must also be considered whether there is a need to specify the temporal performance over time (periodicity) and any other conditions. For example, will these reductions need to be achieved for 85% of the time for a system that is used year-round in a northern US climate with quarterly maintenance visits?

Therefore, the decision endpoint must be specific, and the conditions need to be determined, such as a treatment level to be met (for example, 45 mg/L, 30 mg/L, 10 mg/L) or a performance to be met (for example, allowing a 50% reduction in drainfield trench bottom area without negatively affecting failure rate). By clarifying these types of specific conditions as a part of the decision endpoint, it is easier to evaluate datasets submitted regarding this decision. As many of these specifics as needed can be included in the decision endpoint to help assess the technology performance based upon existing studies or to plan new research studies.

For these purposes, the decision endpoint was revised and the wish to "determine if pretreatment units (three types) will meet BOD_5 and TSS concentrations of 30 mg/L in field systems in the long term" has been specified. This is the new revised decision endpoint of this demonstration for assessment using the DSS.

Step 3: Quantitatively Ranking Each of Ten Different Types of Studies (for example, datasets)

Once a decision endpoint is described, measurement endpoints for datasets will be constructed that are useful for assessing that particular decision endpoint. Measurement endpoints are individual research studies or datasets. They are used to assess whether that decision endpoint has been reached. The available data for making a decision typically would include different types of datasets. As discussed earlier, each measurement endpoint, or research study, has its own unique value for assessing the decision endpoint. Essentially a measurement endpoint is a study or dataset submitted for quantitative scoring and consideration. This is a different approach than used in the Mass DEP (1995) document. Since, for these purposes, measurement endpoints are studies, it is appropriate to assess the weight for different types of studies (that is, different types of datasets). This can, and should, be done prior to evaluation of the actual data included in a dataset from a particular study. By doing so, this provides clear guidance to the manufacturer (or other organization that is requesting an approval) on the relative weight that will be given by the expert panel to different types of studies.

One measurement endpoint might be that BOD_5 and TSS concentrations below 30 mg/L in a *test center dataset* such as an NSF or ETV study indicate that the system will function at that level for the long term. Another measurement endpoint might be that BOD_5 and TSS concentrations below 30 mg/L in a *field study conducted by a third party* indicate that the system will function at that level for the long term. Another measurement endpoint might be that BOD_5 and TSS concentrations below 30 mg/L in *manufacturer-developed datasets* indicate that the system will function at that level for the long term.

As one can surmise, there would likely be different weights assigned to each of these datasets (for example, those in Table 3-1). However, there is a critical point to consider: The rank or numerical value assigned to a study is always conditional upon the decision endpoint. For instance, looking back at Table 3-4, it is obvious that a dataset that is highly valued for one decision endpoint might have much less value for a different decision endpoint.

The decision endpoint for this DSS demonstration is specifically to "determine if pretreatment units will meet BOD_5 and TSS concentrations of 30 mg/L in field systems in the long term." Each of the study types in Table 3-1 can be numerically ranked relative to its applicability for this decision using the expert panel approach. It is recommended that this expert panel have at least five to ten members, and a national or regional organization such as NOWRA, or SORA, for example, could certainly be helpful in assembling these experts, particularly with panels of ten or more. For illustration purposes only five expert panel members are shown in Table 3-5.

Table 3-5
Ranking of Study Types for the Decision Endpoint "Determine if Pretreatment Units Will Meet BOD_5 and TSS Concentrations of 30 mg/L in Field Systems in the Long Term"

Study Type from Table 3-1	Expert Panel Member #1	Expert Panel Member #2	Expert Panel Member #3	Expert Panel Member #4	Expert Panel Member #5	Average or Median Ranking as Appropriate
A						
B						
C						
D						
E						
F						
G						
H						
I						
J						

The makeup of the expert panel should primarily include scientists and regulators, but should also include some manufacturer and industry representatives. It should not be dominated by any one group. Each member of the expert panel independently (without undue influence by other panel members) ranks each of the ten study types from 0.1 to 1.0 for that particular decision endpoint. A rank or numerical score of 1.0 is given to the most appropriate study type for that decision endpoint, and a rank of 0.1 is given to the least relative study type with intervening numerical ranks given to the others. These individual ranking scores are entered by each panel member into a copy of Spreadsheet #1 in the DSS.

The rankings from each of the panel members are averaged and the average numerical ranking is used to determine the value for that type of study or dataset. By using an expert panel of approximately ten members, any unusual rankings will not have undue influence on the final average results. Alternatively, it may be appropriate in some instances to use geometric means rather than averages if the panel assessments are not normally distributed.

Step 4: Assigning Numerical Weights for Each of Eight Data Quality/Quantity Attributes

Once the rank for each type of study or dataset is determined, the process of weighting different characteristics of these datasets is conducted, again using the expert panel approach. These weights are assigned prior to evaluating the actual data submitted, and reflect the intrinsic value of different data quality attributes, not the actual data that is collected for a study.

Each person on the expert panel evaluates and assigns numerical weights from one (1.0) to eight (8.0) for each of the eight data attribute properties. These weights are recorded by each panel member on a separate copy of Spreadsheet #2. One (1.0) is the minimum weight that can be assigned for a data attribute that has little value for that particular study type and decision endpoint. Eight (8.0) is the maximum weight that can be assigned for the most valuable attributes. At least one of the data attributes (the most valued one) must be assigned a value of 8.0. Then, weights between 1.0 and 8.0 are assigned as appropriate for each of the other seven data attributes summarized in Table 3-6 and described earlier in more detail in Table 3-2. This process is conducted separately for each of the ten different study types (such as those types of studies or datasets indicated in Table 3-1). Table 3-6 illustrates a ranking score sheet for one of the expert panel members. As in Step 3, the geometric mean or median can be used if necessary.

Once an expert panel member determines his or her assessment of the numerical weights for the eight data attributes for one type of study or dataset, the weights are recalculated on a relative scale of 0.0 to 8.0. The weights are compared to each other within a study type and adjusted on a scale of 0 to 8 so the total cumulative potential score for all eight data attributes equals 8. This is handled automatically by formulas in Spreadsheets 4–11. Steps 1–4 should be completed prior to actual evaluation of the submitted data. That is, these four steps should be completed before evaluating any actual research study or dataset. Hence, the data attribute weights for each type of study are determined by quantitatively assigning a score for each of the eight data quality attributes previously described in Table 3-2. The relative value (rankings) for different types of research studies (for example, datasets) is also determined in advance of the actual data analysis.

Table 3-6
Weighting Sheet for Individual Panel Members to Use for Weighting Eight Data Quality/Quantity Attributes

Study Type	Performance Data	Flow Data	Replication	Experimental Control	Range of Environmental Conditions	O&M Applied During Study	3rd Party Assessment	Peer Review Process
A								
B								
C								
D								
E								
F								
G								
H								
I								
J								

Data Quality and Quantity Attributes From Table 3-2 (see Table 3-2 for details)

This is done for each of the ten study types for the decision endpoint to determine if pretreatment units will meet BOD_5 and TSS concentrations of 30 mg/L in field systems in the long term.

The combination of the study type rankings and data attribute weights can provide numerical guidance regarding the quality and quantity of differing datasets needed for a decision before any data is submitted to the regulatory agency.

In other words, since the process allows one to numerically compare the value for different data attributes and study types, the amount and quality of data needed from different types of research studies can be estimated in advance (prior to conducting the studies and collecting the datasets).

Basic numerical data scores (Step 5) can be estimated in advance, and this process can help distinguish between the value of data quantity and quality attributes. This prior ranking and weighting can then be used to guide the manufacturer regarding how much data of what quality must be submitted to the regulatory agency before consideration of an approval is appropriate. The manufacturer that will be requesting a regulatory decision can also use this as a guide in deciding what types of studies and data attributes are of most value and worthy of the company's investment of scientific and research resources. Finally, this process allows the regulatory agency to determine how high the bar is for consideration of an approval and possibly to deflect the political pressures that are always present to issue an approval on the basis of limited research.

It is worth noting that Steps 1 through 4 of the DSS process may well identify where there is a lack of consistency or agreement among expert panel members regarding the ranking values assigned for different datasets or weights assigned for data quality attributes. Conducting these steps prior to data collection can help the expert panel clarify its decision-making process and examine its process to determine if biases exist.

Step 5: Determining the Value or Data Score for the Research Study Data

The next step in the DSS process is to examine the research studies (datasets) that have been submitted and to evaluate the characteristics of the data collected during those studies. The relative ranking and value of different study types have already been determined for that decision endpoint. Similarly, the weights for each data quality/quantity attribute have been assigned prior to this point in the process.

What is left to do is to evaluate the data collected and determine the data scores for each of the eight data attributes. These "data scores" are different from the "numerical data weights" previously determined. The data score assigned in Step 5 for each of the data attributes assesses whether the data supports the decision endpoint. The data from each individual study is assessed independently of other datasets, since each study is a measurement endpoint. Even studies that are of the same type (for example, one of the ten study types in Table 3-1) are separately evaluated.

The expert panel evaluates the data from each study separately. First, each panel member reviews the data submitted for that study and independently assigns a data score from 0.0 to 1.0 for each of the eight data attributes. This is done on a separate copy of Spreadsheet #3 by each panel member for each research study. The scores assigned in Step 5 are the panel's assessment of how well the specific data submitted supports (or does not support) the decision endpoint.

If the data are of high quality and strongly support the decision endpoint, then a score of 1.0 is assigned for that data attribute. For example, use of excellent sampling methods and a good Quality Assurance Project Plan (QAPP) during the research study might result in a score of 1.0 for experimental control. Performance data indicating that the technology is performing correctly (for example, BOD < 30 mg/L) might result in a performance data score of 1.0. If the flow rates tested in the research match well with the flow rates that will be used in the field in real life, then a flow data score of 1.0 would be assigned.

However, if the data submitted are of low quality (for a particular data attribute) or if the data do not support the decision endpoint, then that data attribute receives a score of 0.0. Examples include measured performance data showing an unacceptably high BOD level compared to the decision endpoint level, samples being poorly collected using the wrong technique, flow data for the test systems not matching well with the expected use, and loading rates tested not matching the proposed loading rates in the decision endpoint.

Each of the following eight data attributes in the study are considered and scored separately from the first attribute (performance data) to the last attribute (level of peer review used in the study).

1. Performance data (for example, BOD and TSS removal measurements for this demonstration) collected during the study are compared to the decision endpoint.

2. Flow rates used in the study tests are compared to flow rates that will be used in real-life systems.

3. The degree of replication used in the study is evaluated.

4. The extent of experimental control used during the study is assessed.

5. The range of environmental conditions tested during the study is compared to the range of environmental conditions—rainfall, temperatures, soil conditions—that usually occurs in the regulatory jurisdiction where the technology is proposed for use.

6. The O&M used during tests of the technology in the study are compared to the level of O&M proposed for real-life use of the technology.

7. The independence of the agency or individual collecting the samples, evaluating the data, and preparing the project report is assessed to determine the extent of third-party value.

8. The degree of peer review used during design, implementation, and development of the project report is assessed and scored.

Each of these data attributes is scored by the expert panel members in their independent assessments of the data submitted. These are recorded on separate copies of Spreadsheet #3, and the average data scores for each of the data attribute categories are determined. This is done automatically by formulas in Spreadsheet #14.

Step 6: Summing the Results of the Calculations

The next step of the DSS process is to quantitatively calculate the results of the expert panel evaluations. During each stage of the process the averages for each individual expert panel score have been determined. These are included in Spreadsheets #12 through #14. The numerical scores (average scores) are then used in Spreadsheet #15 for calculating the overall decision endpoint score. This score gives guidance to the regulatory community regarding a scientifically-based decision concerning a proposed regulatory approval. The equation used is:

> Y = Sum from all studies of (ranking A × (sum of (data attribute relative weight B × data score C for each attribute))),

where:

A = the numerical rank value for a measurement endpoint (that is, a particular study type),
B = the calculated relative weights for the eight data quality/quantity attributes, and
C = the data scores assigned for the eight data attributes.

Note that:

- *A*, the study type numerical rankings, can range from 0.1 to 1.0.
- *B*, the data weights, can range from 1.0 to 8.0.
- *C*, the data scores, can range from 0.0 to 1.0.
- All of these are relative to the decision endpoint being considered.

The maximum score obtainable using this system for any one study is a score of 8.0; the minimum score is 0.0. The score for each study is calculated separately. Then, for all studies (measurement endpoints) that are pertinent to the decision endpoint being considered, the totals are summed. For example, if three studies were conducted and evaluated using the DSS process and the scores were 4 points, 6 points and 1 point, then the total score for the decision endpoint would be 4 + 6 + 1 = 11 points.

Step 7: Comparing the Calculated Scores to a Predetermined Standard

The final total score for all studies is then compared to the previously determined decision criteria, or standard value, needed to recommend approval of a regulatory decision.

The decision levels suggested earlier are reiterated here:

- Piloting/Testing Confirmation Use – 6 points
- Extended Innovative Product Use – 12 points
- General Accepted Use – 24 points

Using this guidance, the hypothetical scores just given (11 total points) would support a regulatory decision for use of a technology at a piloting stage. This would allow data collection and assessment, but not broad use of the technology.

Summary of Decision Support System (DSS)

Keep in mind that the DSS is a decision support system, not a decision-making system. It is to be used as a guide to bring more science into the decision-making process by incorporating the scientific weight of evidence approach. If the process is used with an eye to minimizing bias and with a large and diverse expert panel that has substantial research, industry, and regulatory expertise, the DSS should greatly assist regulators in making scientifically defensible decisions.

Also, if the DSS process is used prior to assessing datasets (the manner described here), there will be a tangible benefit to manufacturers. It could help them identify important research needs as well as guide them in the value that could be expected from allocating their frequently limited research program resources toward the correct types of research studies (datasets). This will only be the case, however, if the decisions made by the expert panel up front (that is, the numerical study ranking and quantitative data attribute weights assigned for different types of studies) are followed consistently when the actual data is submitted for evaluation.

For the pretreatment technologies decision endpoint used here, it is useful to have guidance regarding the importance of different types of datasets. The previous chapter illustrated a statistical comparison of the value of test center data versus field data for predicting field performance of pretreatment technologies. This assessment is useful for helping guide the expert panel members regarding the weights they assign when using the DSS for this decision endpoint

4 CONCLUSIONS & DISCUSSION

Everyone involved in advanced onsite wastewater treatment decision-making is looking for different things. The regulator wants to make sure that the technology he or she approves will perform as claimed. The manufacturer wants to get an approval for use and also wants to have some certainty about the approval process—what it will cost, how much time it will take, and what types and levels of research effort are needed to obtain an approval. The homeowner wants a solution that has been properly reviewed and approved to ensure performance at minimum cost.

This research project had a number of major objectives as stated in Chapter 1. A summary of how the project addressed each of the objectives along with pertinent conclusions follows:

Assemble Valid Quality Test Center and Field Data Into Unified Sets and Evaluate Their Relative Qualities

Biochemical Oxygen Demand (BOD) and Total Suspended Solids (TSS) data from National Sanitation Foundation (NSF) International Standard 40 evaluations, Environmental Technology Verification (ETV) projects, National Onsite Demonstration Projects (NODP), and data collected by regulatory agencies and vendors was assembled and reviewed to eliminate duplicate samples, samples from non-residential facilities, and others.

Analyze the Datasets Statistically to Prove or Disprove the Null Hypothesis if Test Center and Field Data Distributions Are Similar or Dissimilar

If data distributions are similar, then predict field performance relationships. If data distributions are dissimilar, then develop the best possible fit for these relationships.

The statistical analysis concluded that the variability associated with test center data was significantly less than the variability of data collected from real world situations. Further, the analysis leads to the conclusion that to best predict the variability of a technology’s performance, 18 to 37 systems should be sampled on a random basis 4 times over the course of 1 year. The exact number of facilities and samples is dependent on the level of confidence desired by the reviewer (such as 18 for 95%, 37 for 99%). The degree of confidence in this study did not increase significantly as the number of facilities or sampling frequency was increased. The validity of these conclusions was not evaluated for parameters other than BOD and TSS.

Develop a Decision Support System (DSS) for Ranking or Weighting Different Types of Data

The DSS guides regulators and manufacturers regarding the possible combinations of test center and field data needed to allow state/county/local approvals of new technology as "proven."

The concept of scientific weight of evidence was demonstrated using numerous examples and led into the development of the Decision Support System (DSS). The DSS was described, documented and demonstrated in the report and accompanying CD Resource Tool. The DSS can be used for the evaluation of any individual technology for any particular parameter of interest.

The DSS can be used by as few as two or three individual experts but has increasing acceptability with increased numbers of reviewers and increasing breadth of their experience. If a small expert panel is used, at least one of the expert panel members should be a scientist with experience in the onsite wastewater field. Larger expert panels should include more scientists, but should also include experienced field onsite wastewater practitioners. Obviously, regulators are critical to this process and should be well-represented in expert panels.

Allow for Greater Acceptance of the National Onsite Wastewater Recycling Association (NOWRA) Model Code

The methodologies developed give increased assurances that management of onsite treatment technology can be successful while still being cost effective. Instead of every jurisdiction having to sample a large number of systems on an extensive basis, a protocol is presented that will allow regulatory and management entities to have confidence in the performance of improved technologies. Additionally, the DSS developed here fits well into the technology assessment approach proposed by the NOWRA Model Code committee. The DSS, if used by NOWRA in a national technology assessment database, can bring an additional scientific basis to technology assessments.

Build Capacity and Understanding in the Onsite Program Arena

The onsite public arena includes:

- Vendors
- Testing organizations
- State regulators
- Consultants
- Implementing and management agencies
- The Public

Both the statistical and DSS methodologies developed here can build capacity and improve understanding for those involved in onsite wastewater treatment by providing clear protocols, which, if performed as prescribed, can speed the development and deployment of cost-effective and improved treatment technologies. The savings of time and money as well as confidence in the expectations surrounding technology performance will improve acceptance of technology by all involved.

Provide a CD Resource Tool

Instruction on the collection, assembly, analysis, and use (weighting and ranking) of data collected at test centers and in the field gives regulators confidence in the predictable performance of new onsite technology.

The CD Resource Tool contains this report and provides interested parties with well-defined and user-friendly templates to assist in performing assessments and evaluations. The number of spreadsheets required in the DSS for sheets 1, 2, and 3 will vary depending upon the number of experts included in the expert panel and the number of studies or datasets submitted for analysis when conducting a particular technology evaluation. The remainder of the DSS is self-calculating following insertion of the values from sheets 1, 2, and 3 into the appropriate remaining sheets. Ample instructions, numerous examples, tutorial PowerPoint file, and extensive interactive pop-up windows within the DSS spreadsheets themselves provide the first time users of the DSS with substantial direction and instruction. Each expert panel, even small panels, should include at least one experienced onsite wastewater scientist.

Conclusions

The evaluation and analysis of the test center and real world (residence) data resulted in a number of findings that most unbiased observers would probably agree with, but also produced a couple of conclusions that are at least mildly surprising. One, the statistical analysis concluded that the variability associated with test center data was significantly less than the variability of data collected from real world situations.

Therefore, the two data distributions are dissimilar and one data distribution set (such as test or field) cannot be used to accurately predict the other. Since the test data distribution cannot predict the field data distribution, one must decide how much field sampling will be appropriate to accurately predict the long-term performance of the technology. This leads to the second conclusion that if you have only a limited amount of time and money, you are probably better off sampling as many sites as possible on a random basis for a few samples rather than thoroughly evaluating a small number of locations for an extended period of time.

Does this mean test center and demonstration project data are not needed? Not really. In order to develop enough data, as many as 40 systems will need to be installed in the backyards of homes and businesses. If there is not an independent, trusted screening process to make sure that any technology installed has a reasonable chance of being successful, 40 or more families or companies will have invested a great deal of money in a totally unproven technology with little or no assurance that what is put in the ground will work.

Test centers and demonstration projects can fill that void. Likewise, field data is also important due to the need for demonstrating a technology's long term, expected field performance by developing a sampling protocol that provides the desired confidence levels. Also, remember, one of the underlying premises of the Decision Support System is that all non-fraudulent data has value. Determining the relative value can be affected by such statistical analyses.

Different states use different processes to decide what goes into the ground. With or without statistical analysis, a technology that does not perform as advertised still leaves some number of homeowners with a problem. While NSF or demonstration projects might fill that bill, each individual state cannot be spoken for. Massachusetts uses a phased approach with extensive evaluation before allowing more than 15 systems to be installed (rather than accepting a short study like NSF Standard 40 on a single system to allow systems). Other states look at a small dataset and then allow unlimited installations with little or no testing or O & M.

The Decision Support System may be a useful way to try to meet all of these expectations. It provides a way to be secure in a decision and as confident as one can be in predicting the future. To date, the DSS has not been applied to any real world cases. The hope was to work with SORA to demonstrate and critique the DSS, but due to the timing of the DSS development, this was not accomplished within the project period. Examples and a tutorial presentation are included in the CD Resource Tool file that accompanies this report. These will assist the regulator (and others) with understanding of the DSS and how to apply it. Real world case application is the next step. Individual states' regulatory agencies are encouraged to work with regulatory, industry, and academic experts inside and outside of their states to apply the methodology. At a minimum, the system allows for the involvement and participation of the best experts that are available; at its best, all involved in the decision-making process can be confident that the best answer was reached.

Finally, while there seems to be promise in applying this methodology to datasets involving the treatment of BOD and TSS, it is not clear that this exact analysis will be applicable to other contaminant treatment technologies, at least not without some modification. For instance, as noted earlier, the effects of temperature on nitrogen removal can be quite marked relative to the effects on BOD and TSS treatment.

It may be necessary when evaluating nitrogen treatment systems to separate data into warm-period datasets and cold-period datasets in order to obtain meaningful relationships. On the other hand, it is possible that the variability associated with what takes place in each home or business that generates wastewater is such that temperature effects can be ignored. Only a statistical analysis will tell.

5 REFERENCES

Hoover, M. T. and D. Beardsley. 2000. “Science and Regulatory Decision Making.” *Small Flows Quarterly, Volume 1, No. 4, Small Flows Forum*, Fall 2000 (Peer-Reviewed Editorial). National Small Flows Clearinghouse, West Virginia University, Morgantown, WV.

Hoover, M. T. and D. Beardsley. 2001. “The Weight of Scientific Evidence.” *Small Flows Quarterly, Volume 2, No. 1, Small Flows Forum*, Winter 2001 (Peer-Reviewed Editorial). National Small Flows Clearinghouse, West Virginia University, Morgantown, WV.

Massachusetts Department of Environmental Protection (Mass DEP). 1995. Draft Report: *A Weight-of-Evidence Approach for Evaluating Ecological Risks.* Prepared by Massachusetts Weight-of-Evidence Workgroup (Included Staff From Mass DEP, US EPA, NOAA CH2M Hill, ENSR, CDM, McLaren/Hart, Univ. of Miami, Menzie-Cura and Associates). www.state.ma.us/dep.

Milliken, G. A. and D. E. Johnson.1984. *Analysis of Messy Data Volume 1: Designed Experiments*. Van Nostrand Reinhold Co.

6 ACRONYMS AND ABBREVIATIONS

ASAE	American Society of Agricultural Engineers
BOD	Biochemical Oxygen Demand
CV	Coefficient of Variation
DEP	Department of Environmental Protection
DSS	Decision Support System
ETI	Environmental Technology Initiative
ETV	Environmental Technology Verification
FAST	Fixed Activated Sludge Treatment
Mass DEP	Massachusetts Department of Environmental Protection
MOU	Memorandum of Understanding
NDWRCDP	National Decentralized Water Resources Capacity Development Project
NEHA	National Environmental Health Association
NEIWPCC	New England Interstate Water Pollution Control Commission
NESC	National Environmental Services Center
NODP	National Onsite Demonstration Project
NOWRA	National Onsite Wastewater Recycling Association
NSF	NSF International (formerly National Sanitation Foundation)
NSFC	National Small Flows Clearinghouse
SORA	State Onsite Regulators' Alliance
TSS	Total Suspended Solids

US EPA	United States Environmental Protection Agency

A FIGURES REFERENCED IN CHAPTER 2 SHOWING VARIABILITY AND RELIABILITY OF TEST CENTER AND FIELD DATA

Figure 2.1.a: Advantex, NSF, BOD Frequency Distribution.

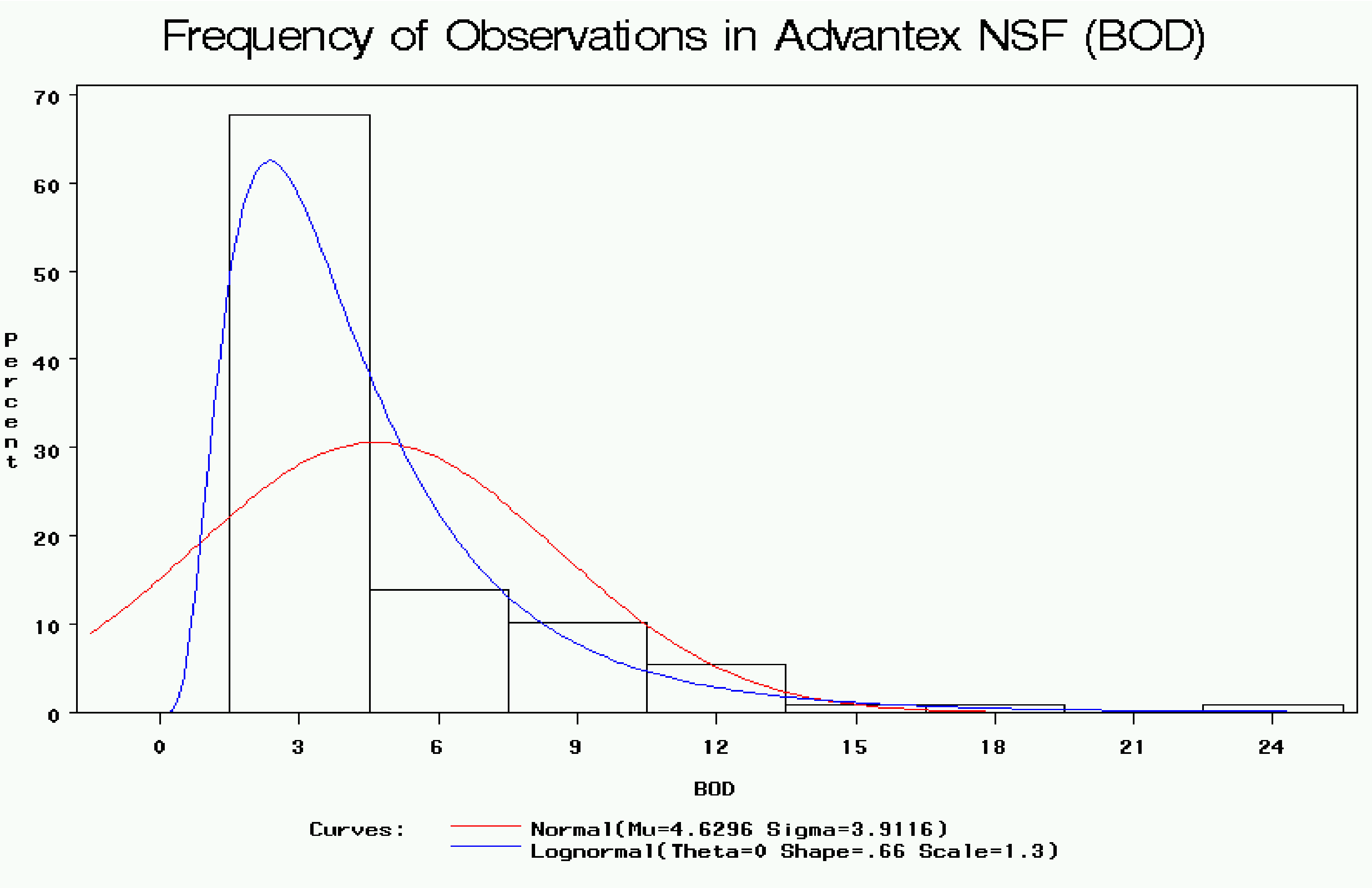


Figure 2.1.b: Advantex, Residences, BOD Frequency Distribution

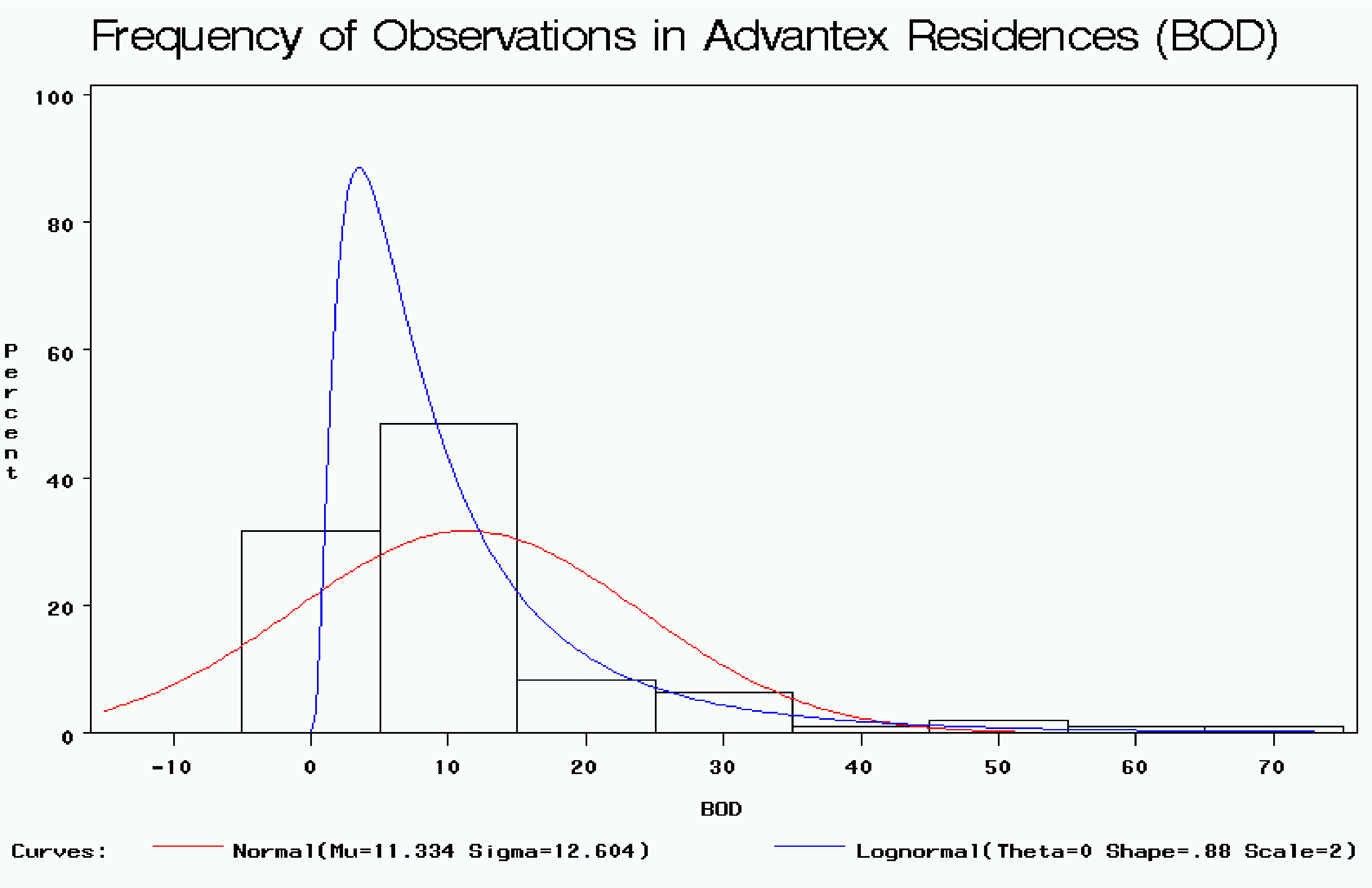


Figure 2.1.c: Advantex, NSF, TSS Frequency Distribution

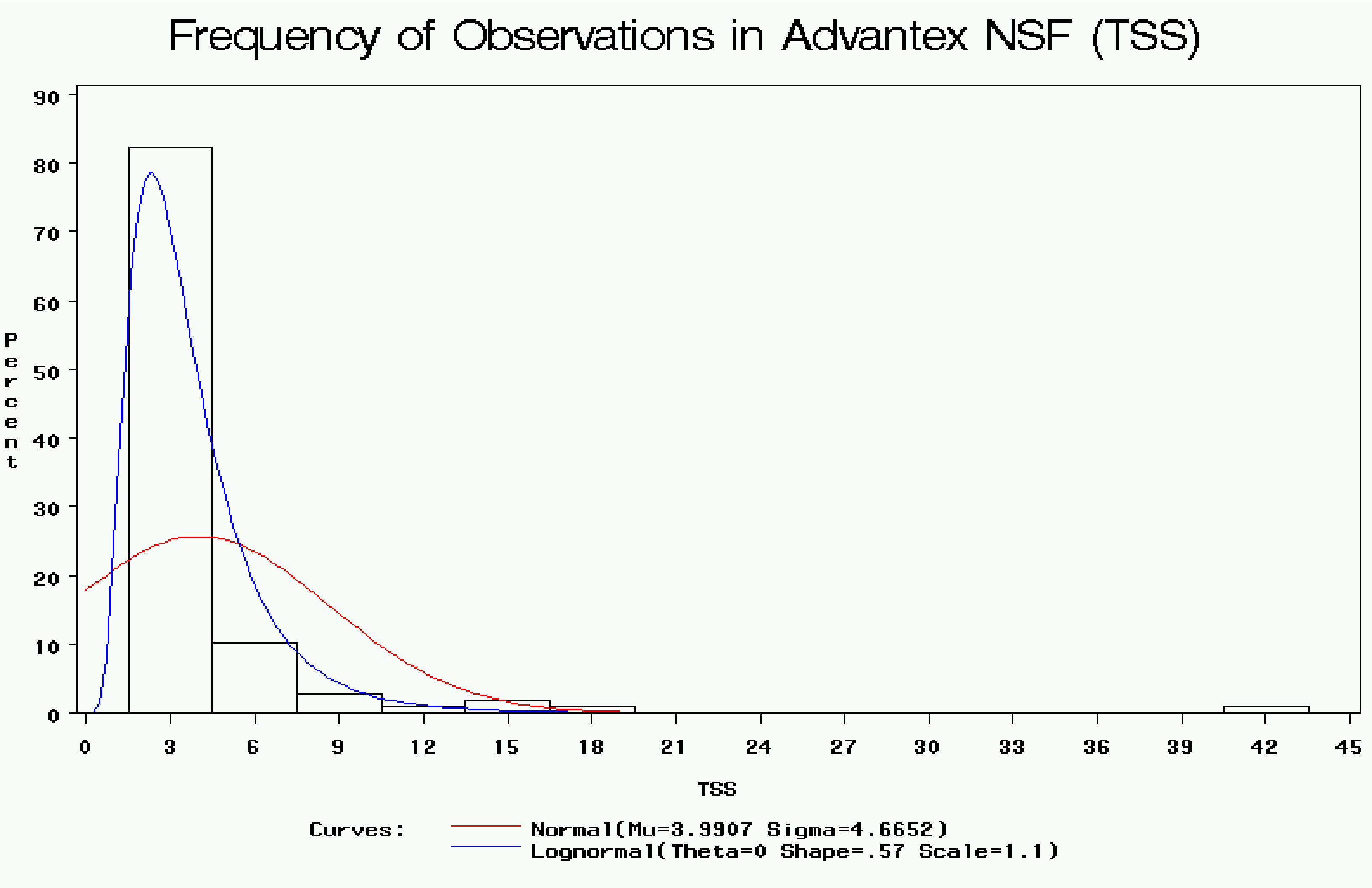


Figure 2.1.d: Advantex, Residences, TSS Frequency Distribution

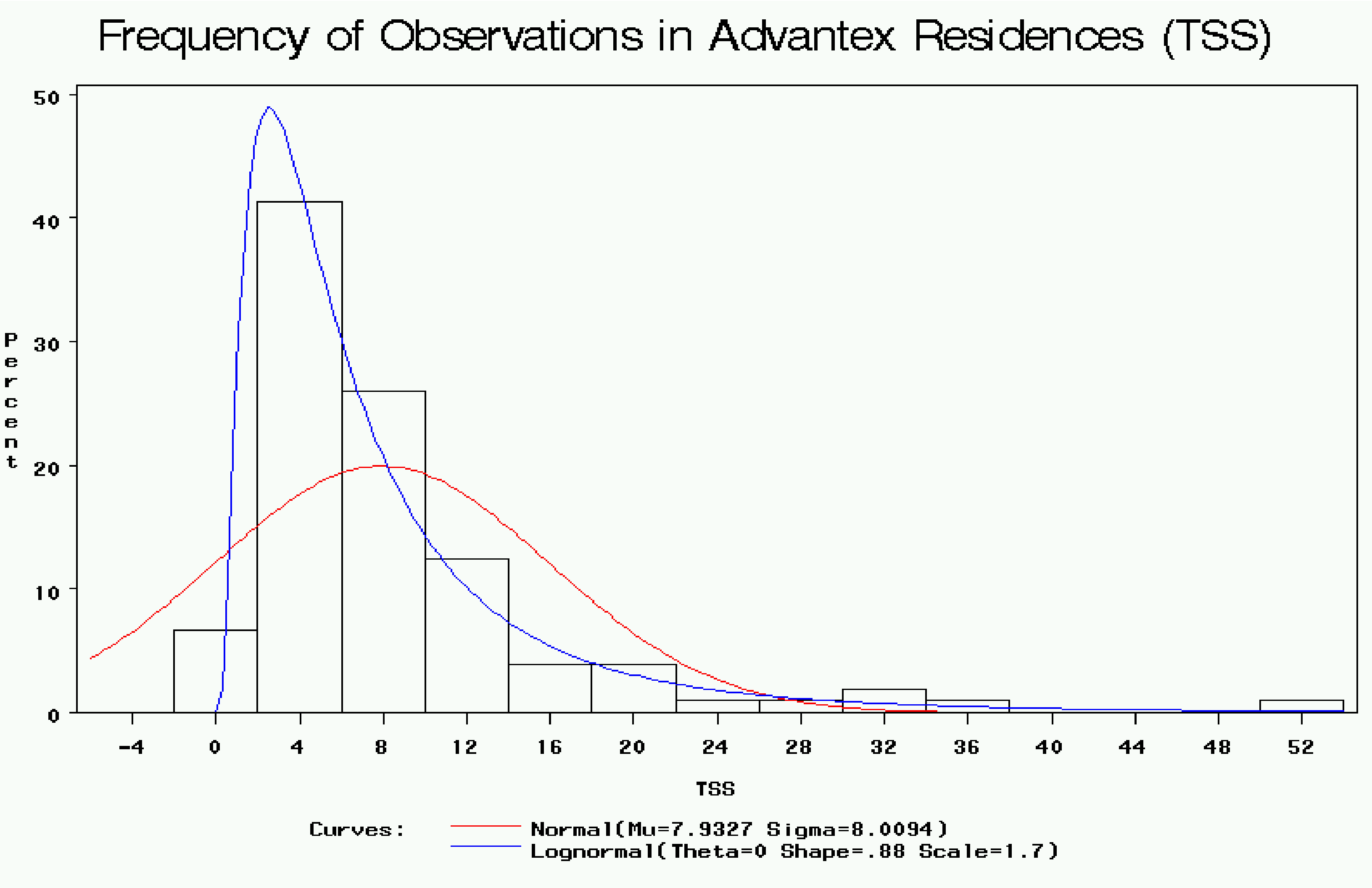


Figure 2.2.a: Bioclere, NSF, BOD Frequency Distribution

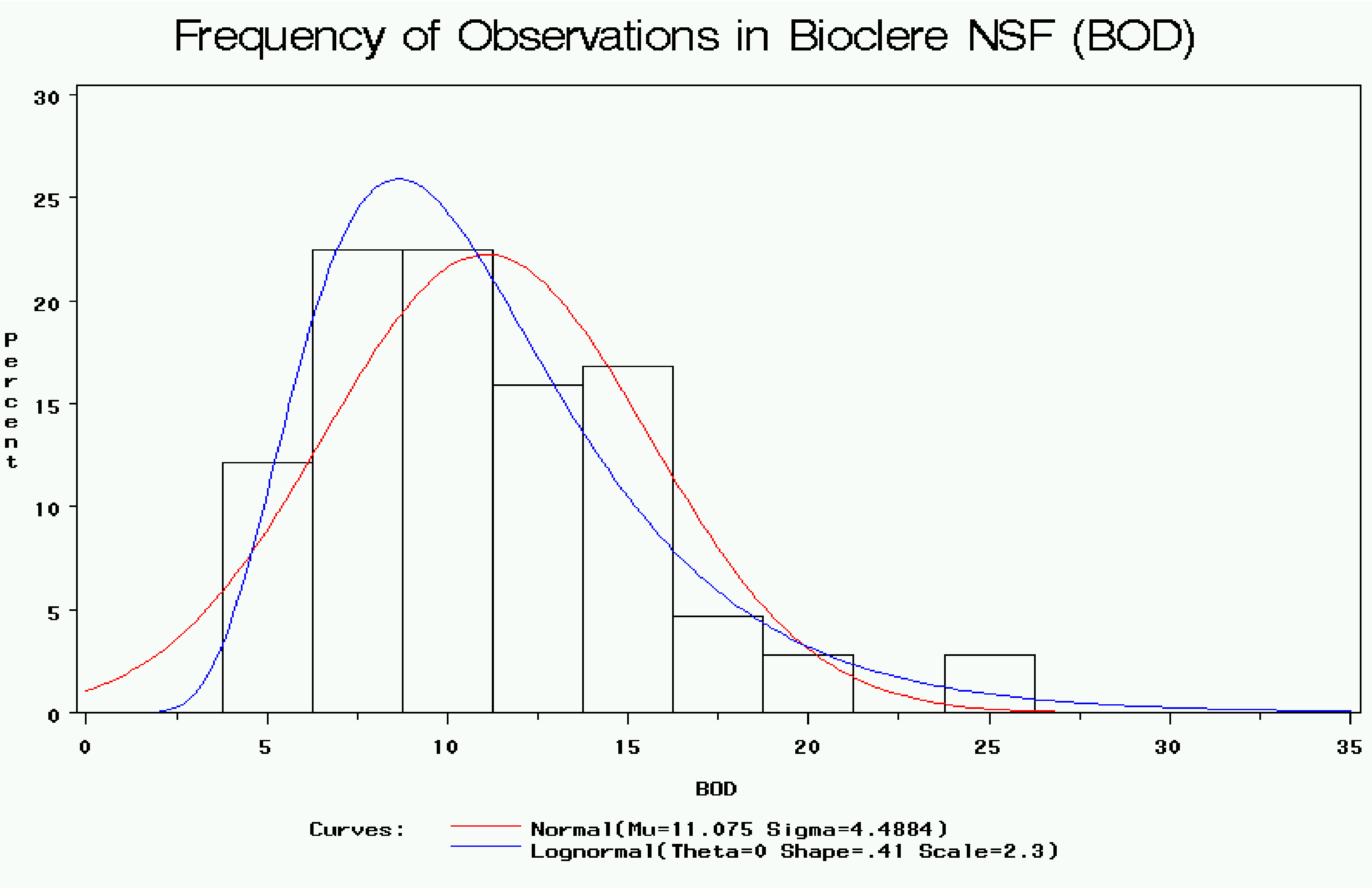


Figure 2.2.b: Bioclere, Residences, BOD Frequency Distribution

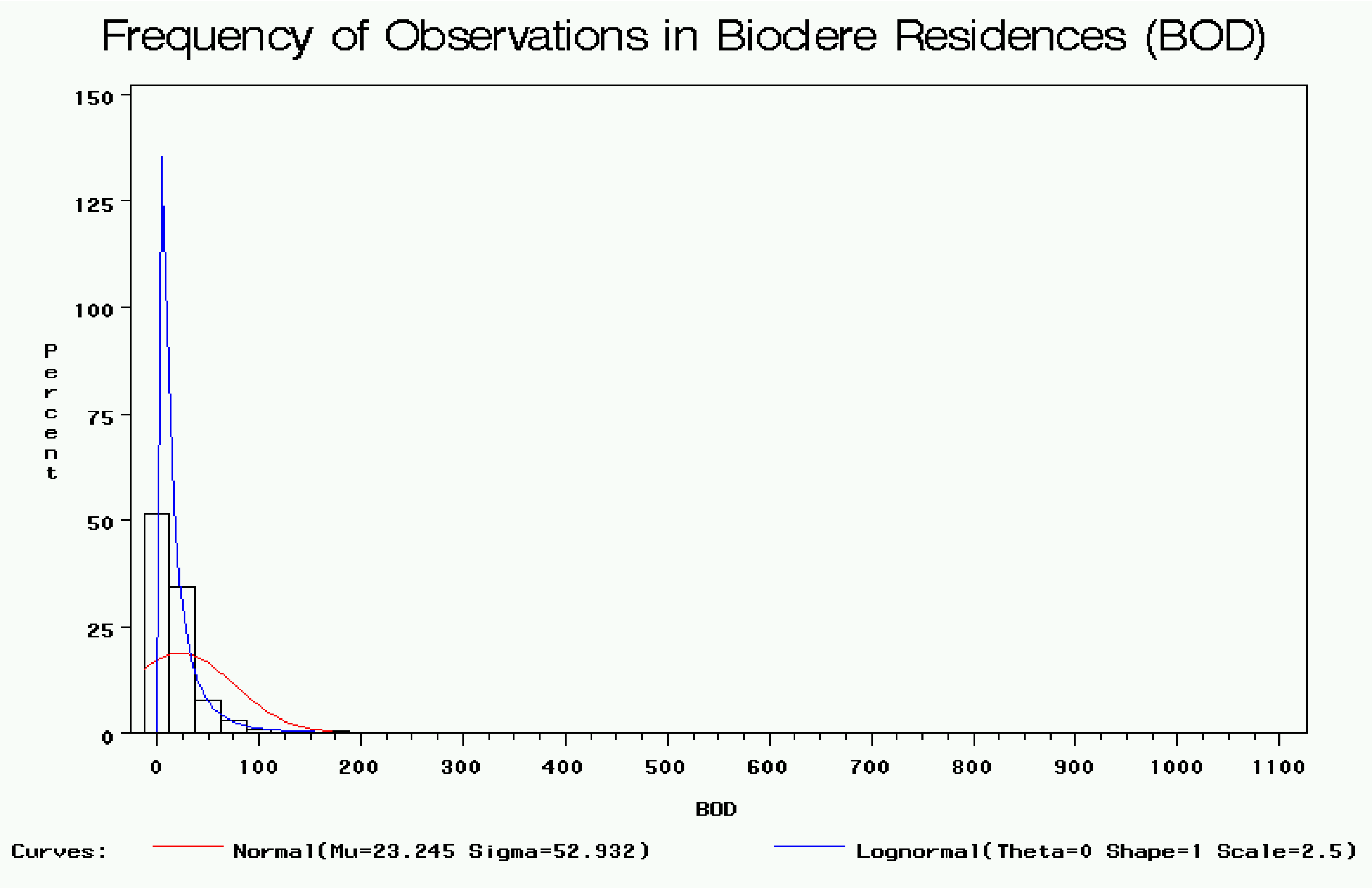


Figure 2.2.c: Bioclere, NSF, TSS Frequency Distribution

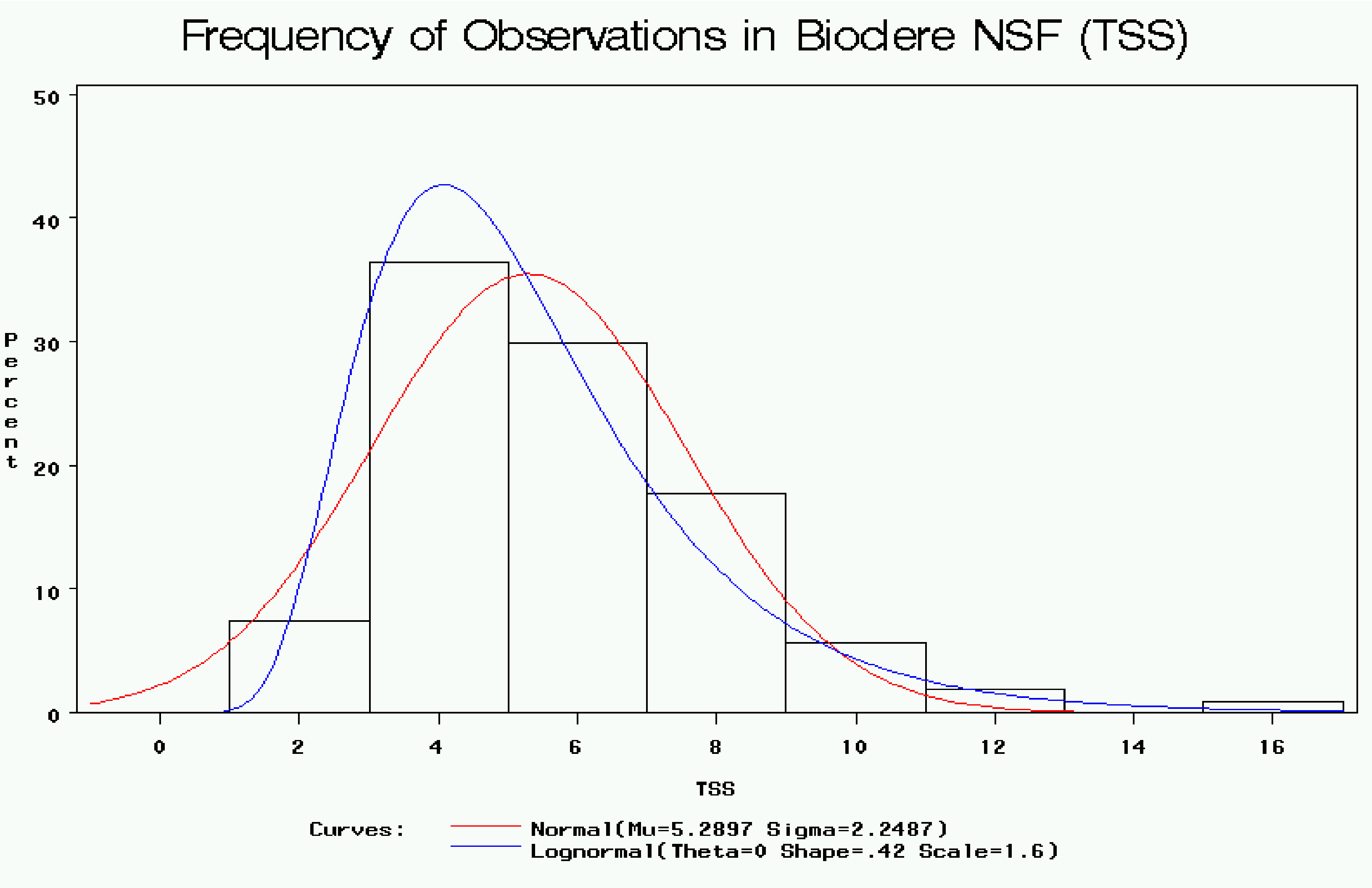


Figure 2.2.d: Bioclere, Residences, TSS Frequency Distribution

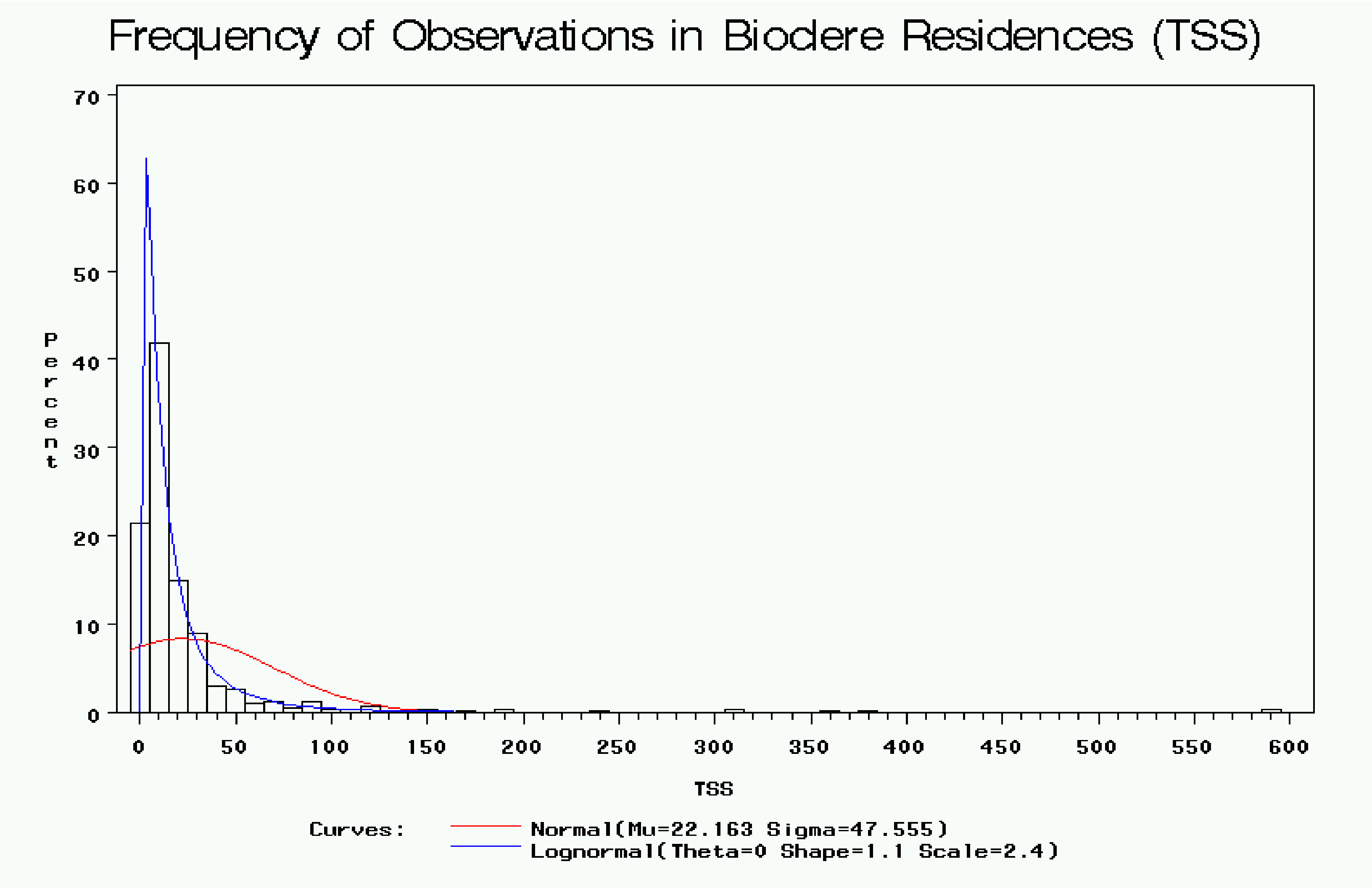


Figure 2.3.a: FAST, NSF, BOD Frequency Distribution

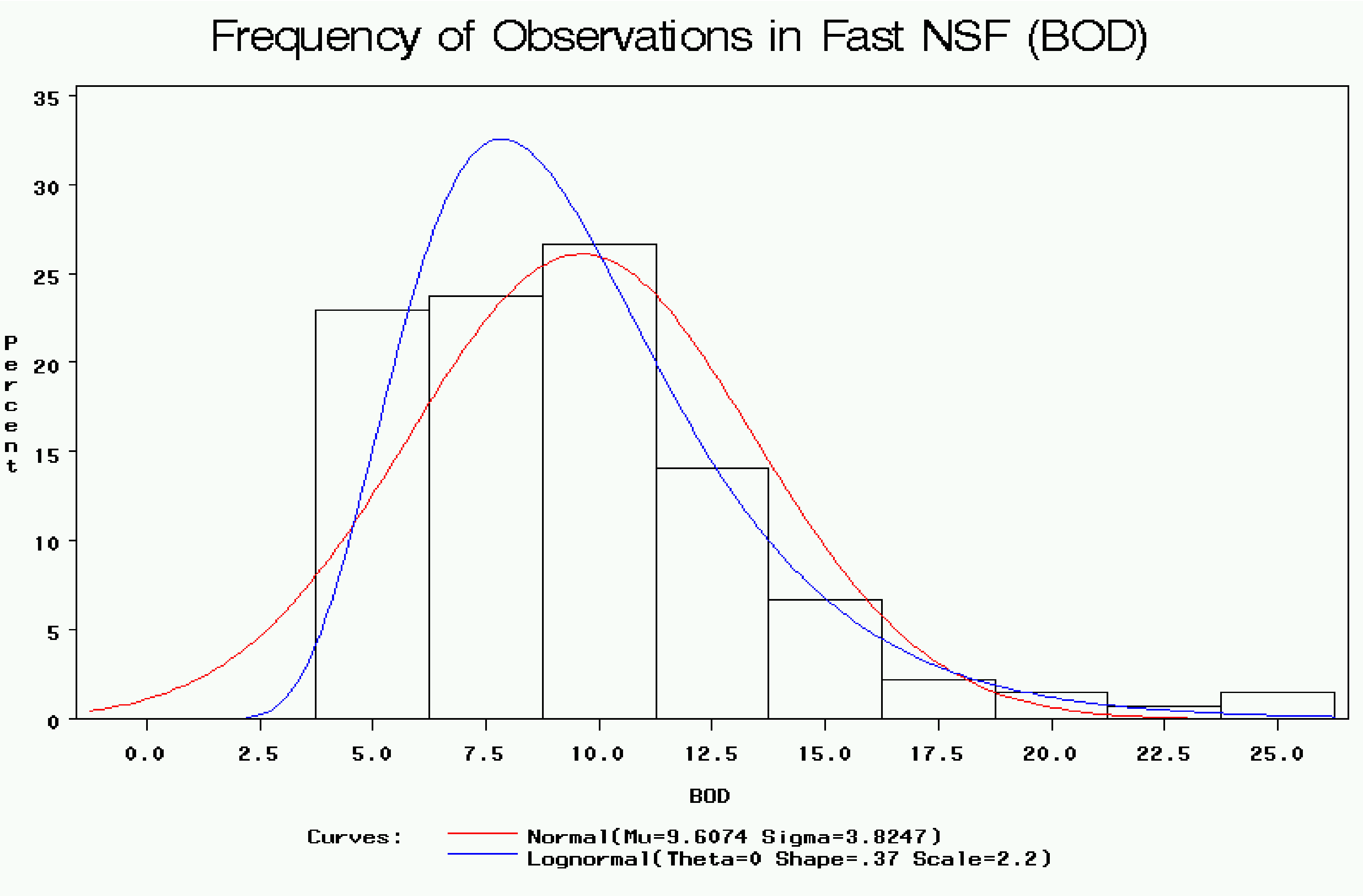


Figure 2.3.b: FAST, Residences, BOD Frequency Distribution

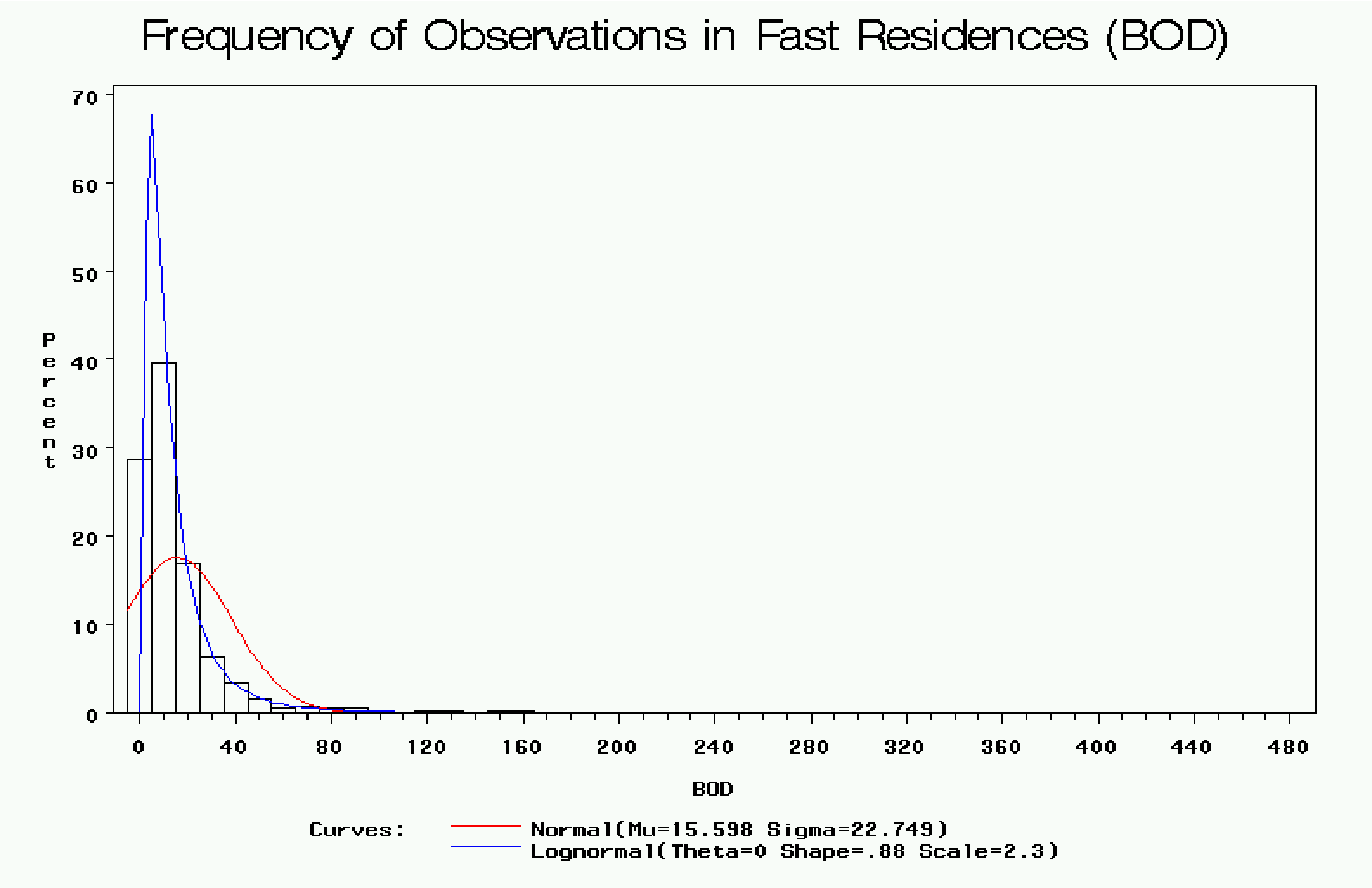


Figure 2.3.c: FAST, NSF, TSS Frequency Distribution

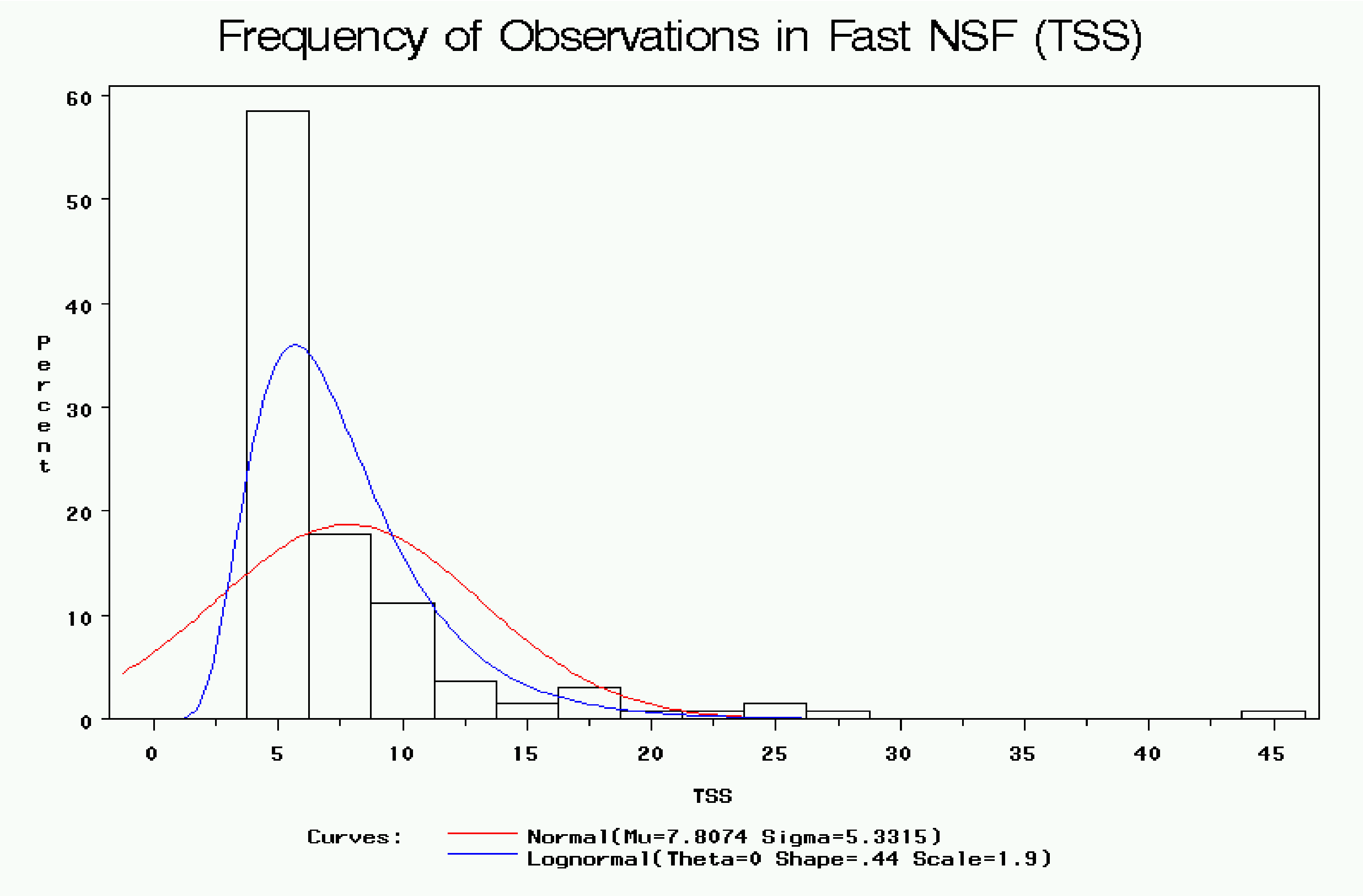


Figure 2.3.d: FAST, Residences, TSS Frequency Distribution

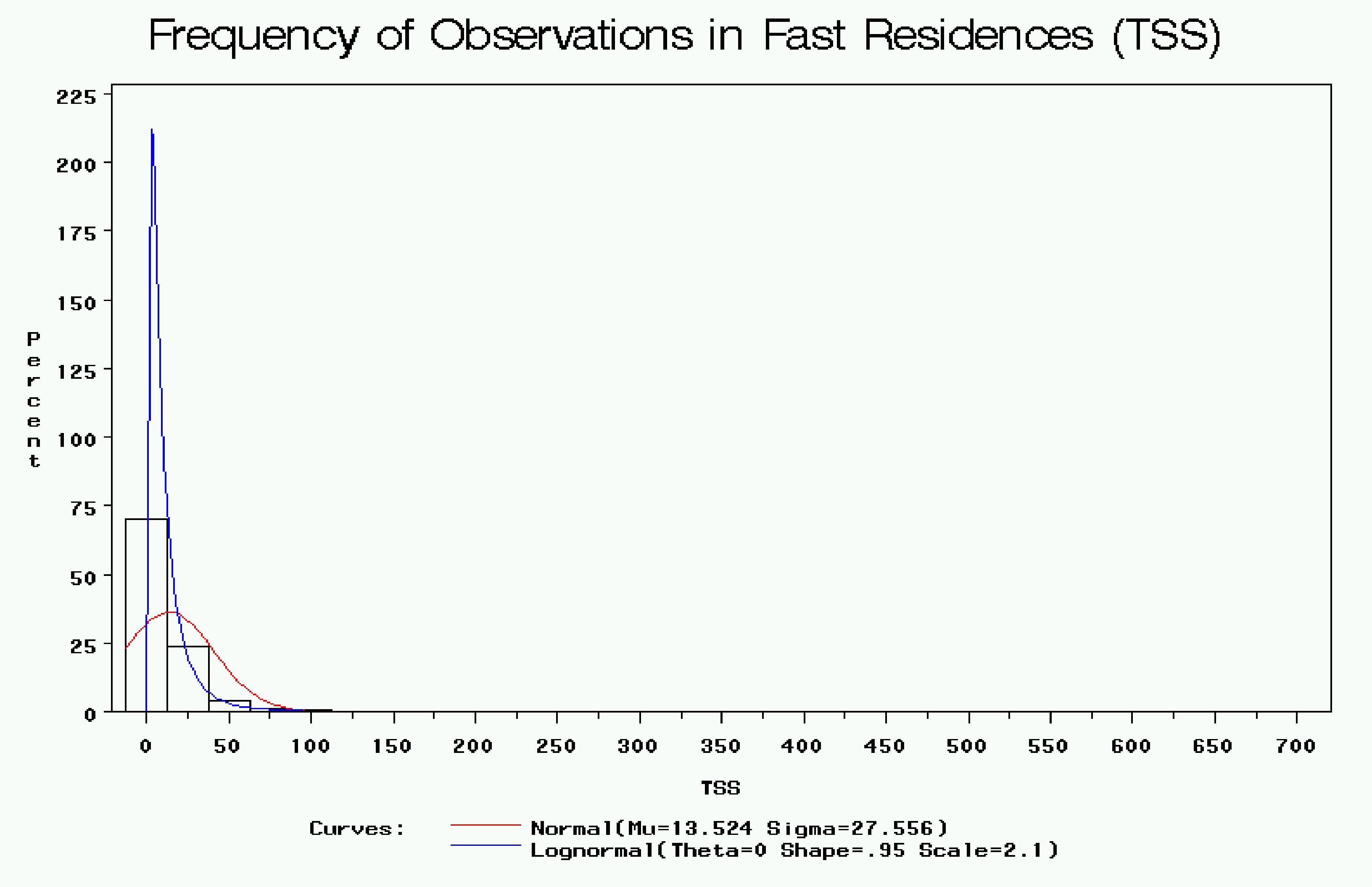


Figure 2.4.a: Distribution of Advantex Log10(TSS) by Site

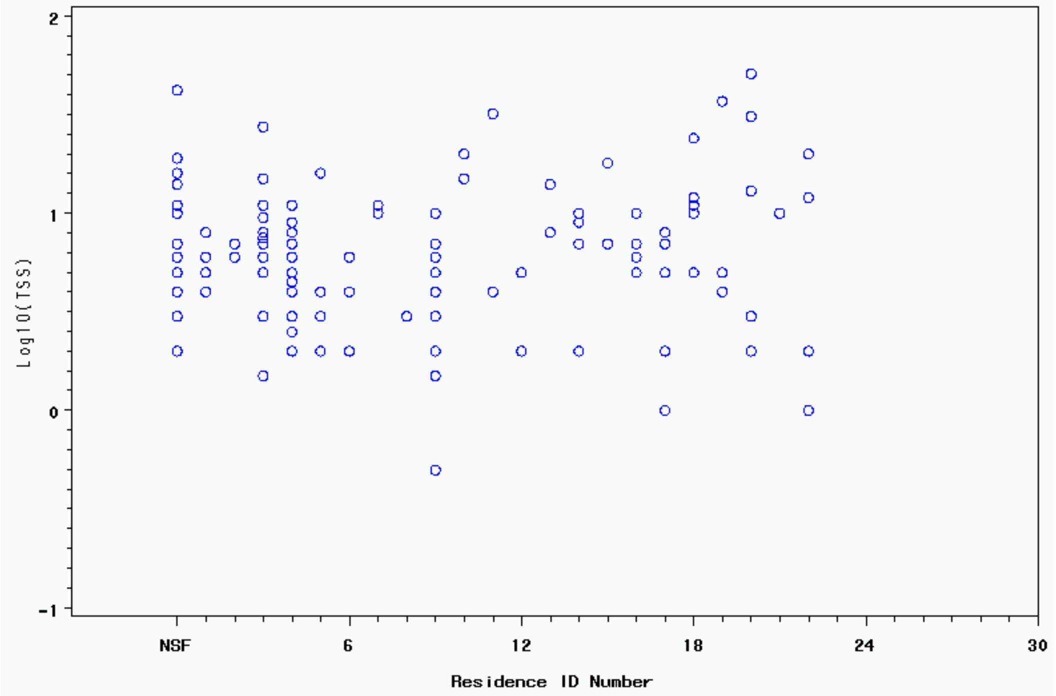


Figure 2.4.b: Distribution of Advantex Log10(BOD) by Site

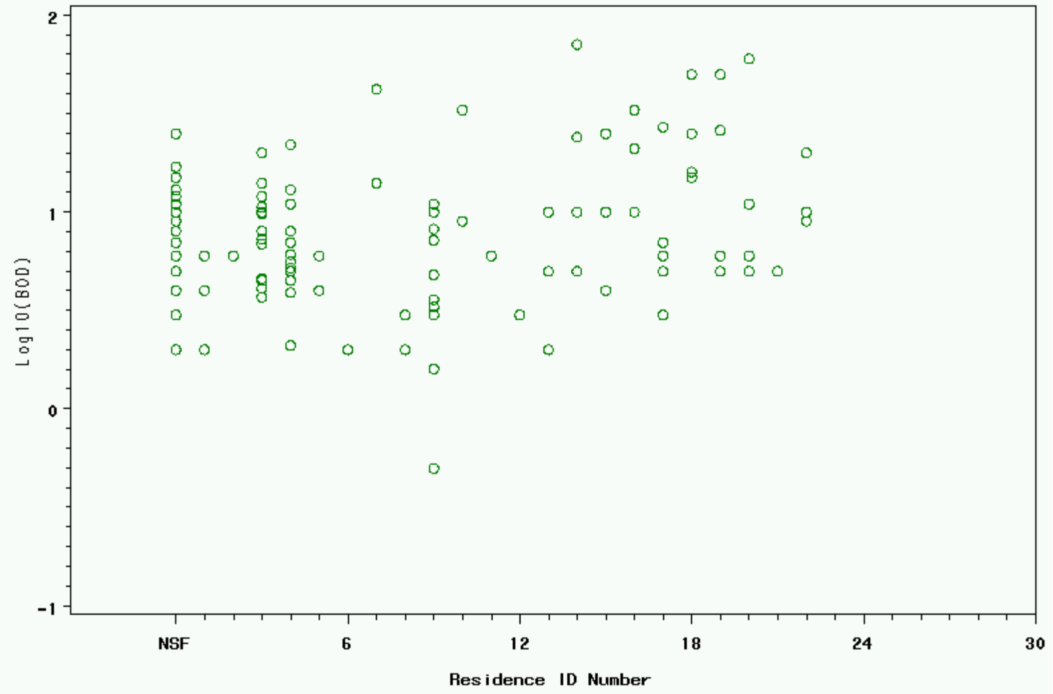


Figure 2.5.a: Distribution of Bioclere Log10(TSS) by Site

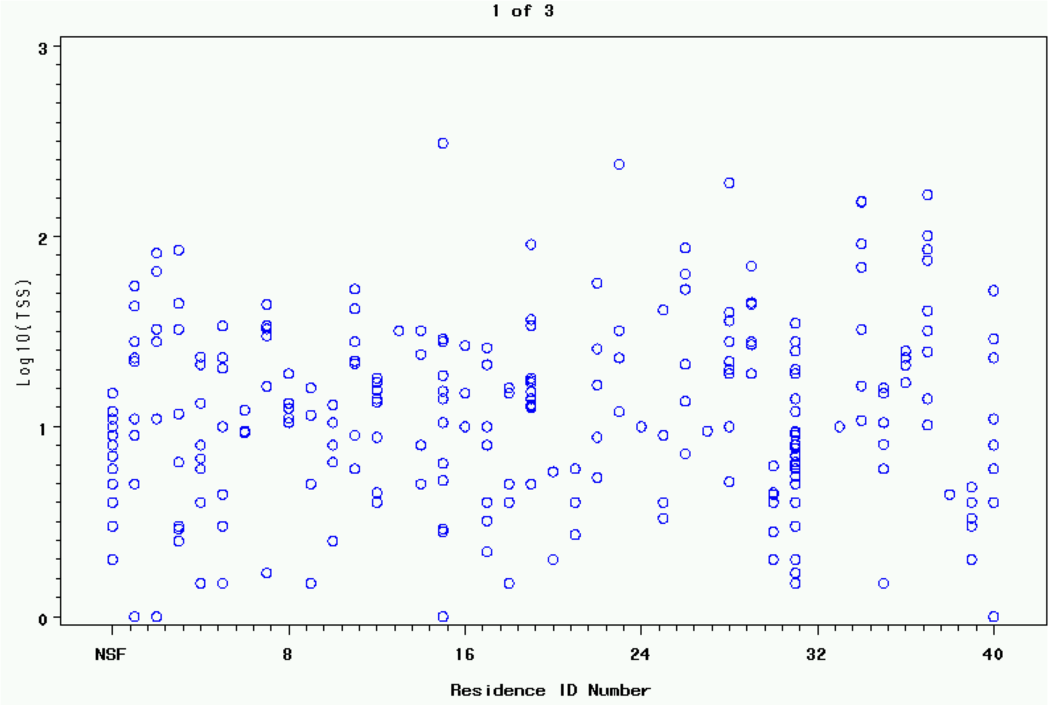


Figure 2.5.b: Distribution of Bioclere Log10(TSS) by Site

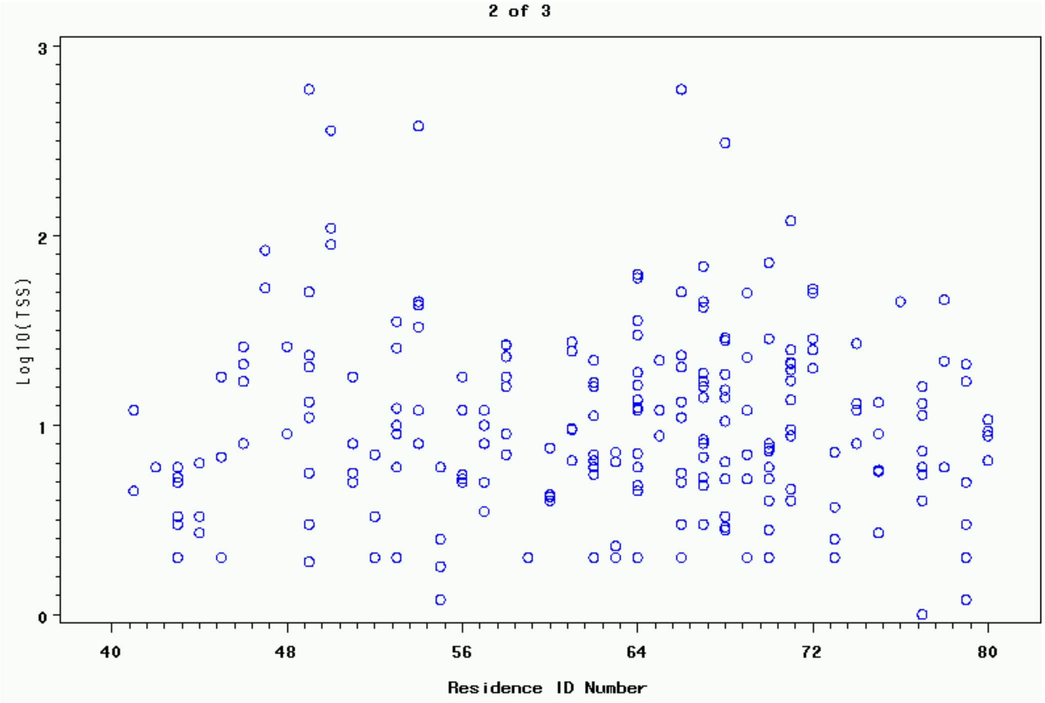


Figure 2.5.c: Distribution of Bioclere Log10(TSS) by Site

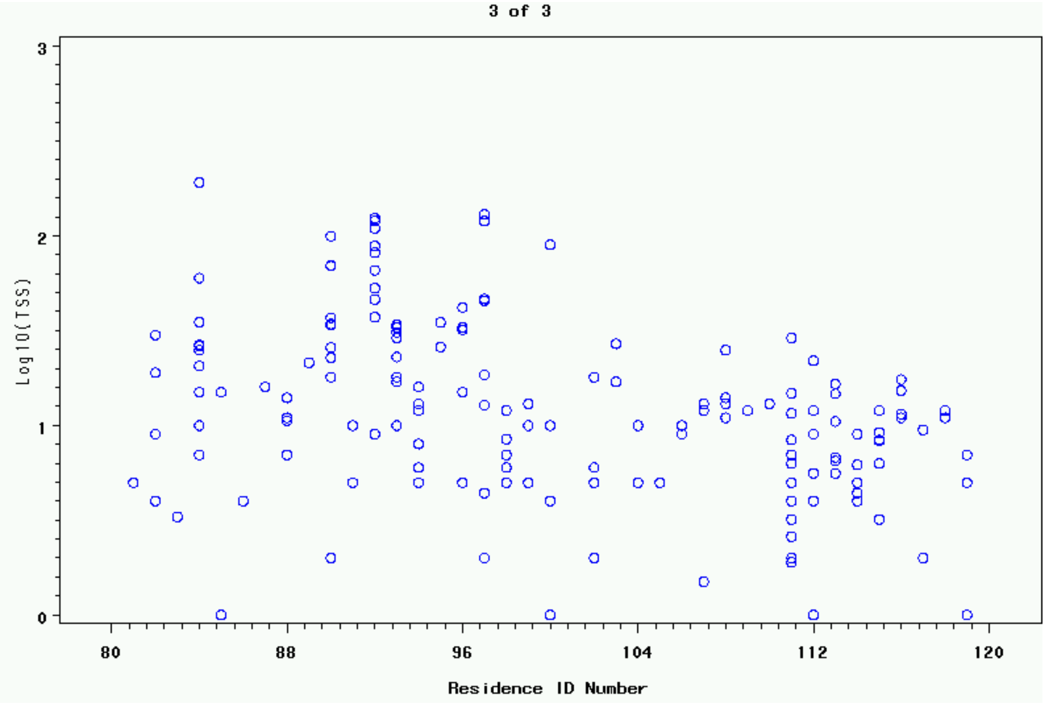


Figure 2.6.a: Distribution of Bioclere Log10(BOD) by Site

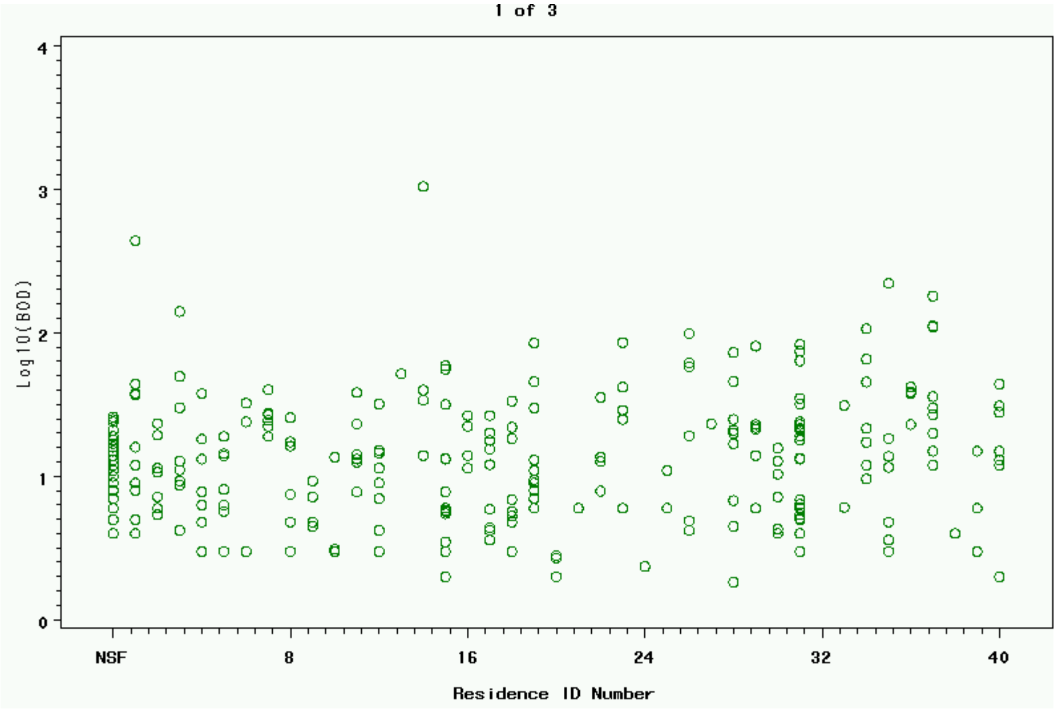


Figure 2.6.b: Distribution of Bioclere Log10(BOD) by Site

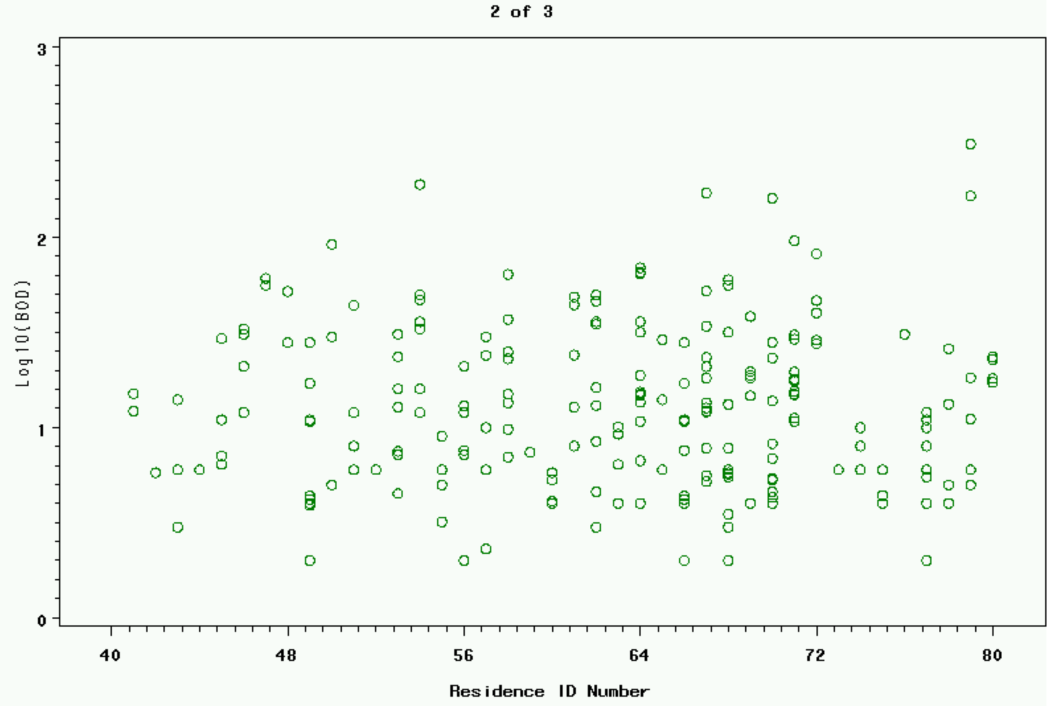


Figure 2.6.c: Distribution of Bioclere Log10(BOD) by Site

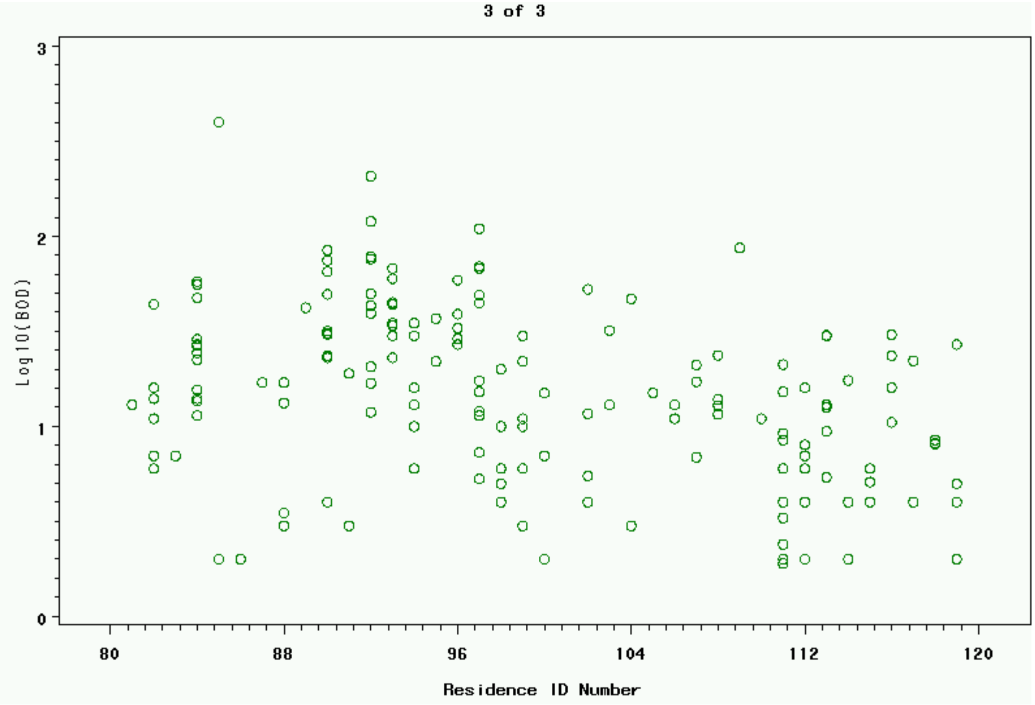


Figure 2.7.a: Distribution of FAST Log10(TSS) by Site

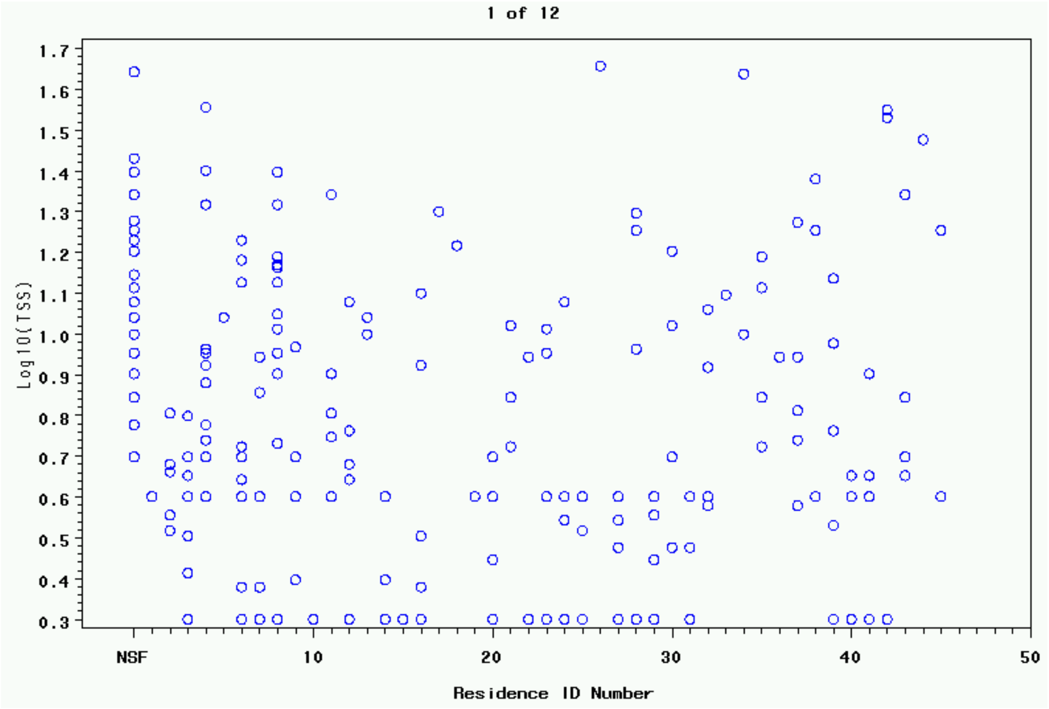


Figure 2.7.b: Distribution of FAST Log10(TSS) by Site

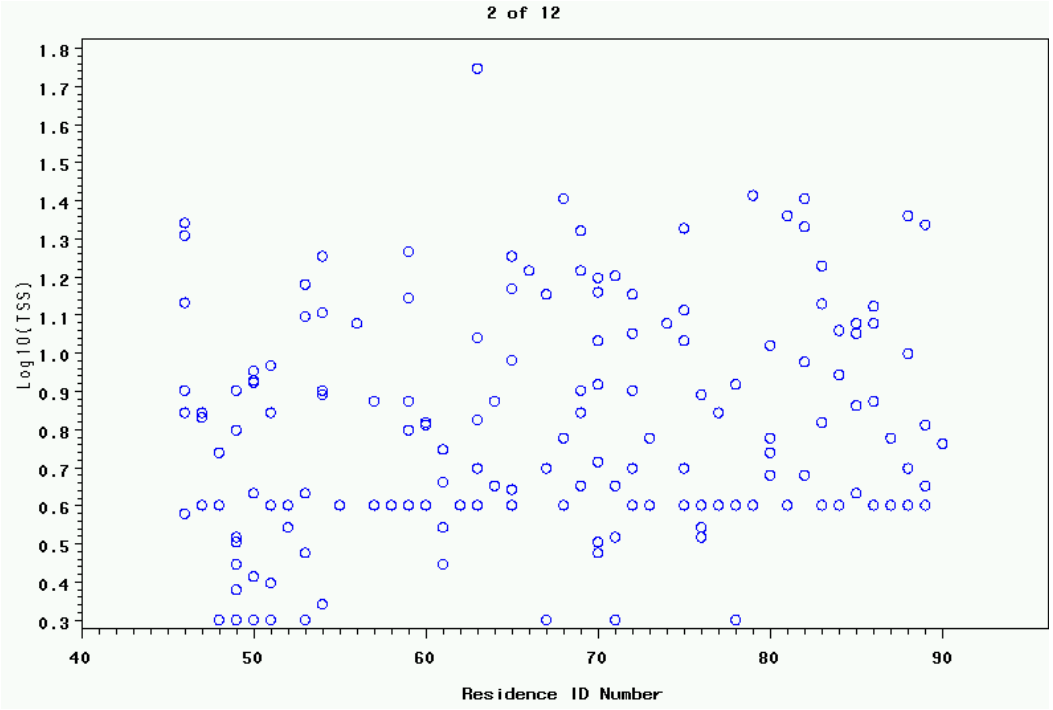


Figure 2.7.c: Distribution of FAST Log10(TSS) by Site

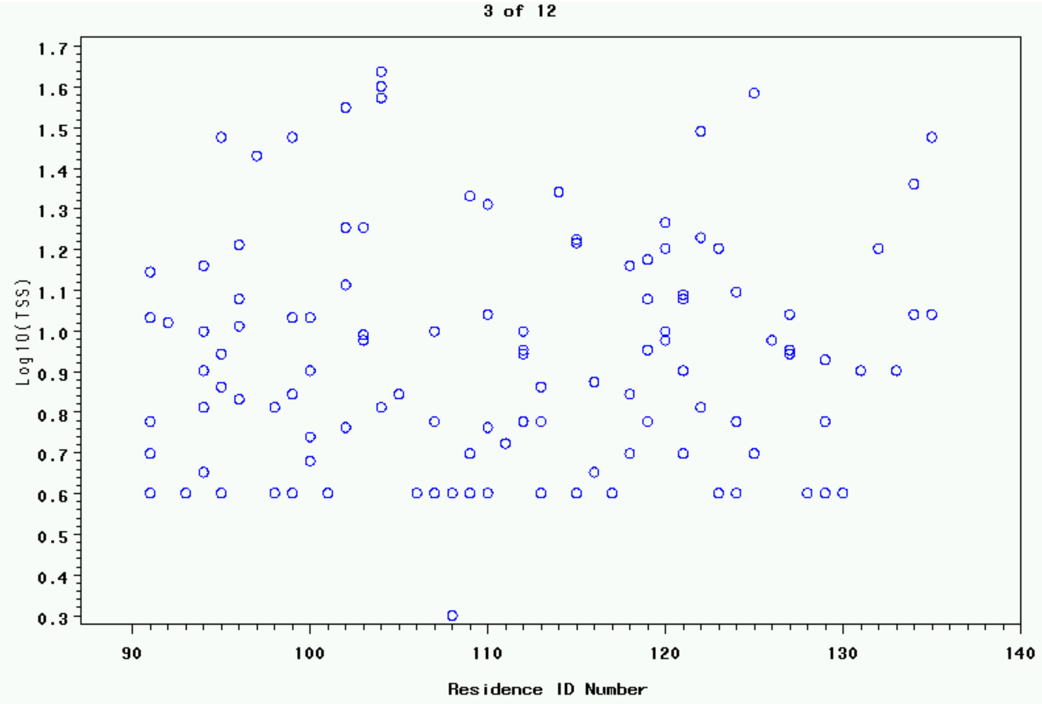


Figure 2.7.d: Distribution of FAST Log10(TSS) by Site

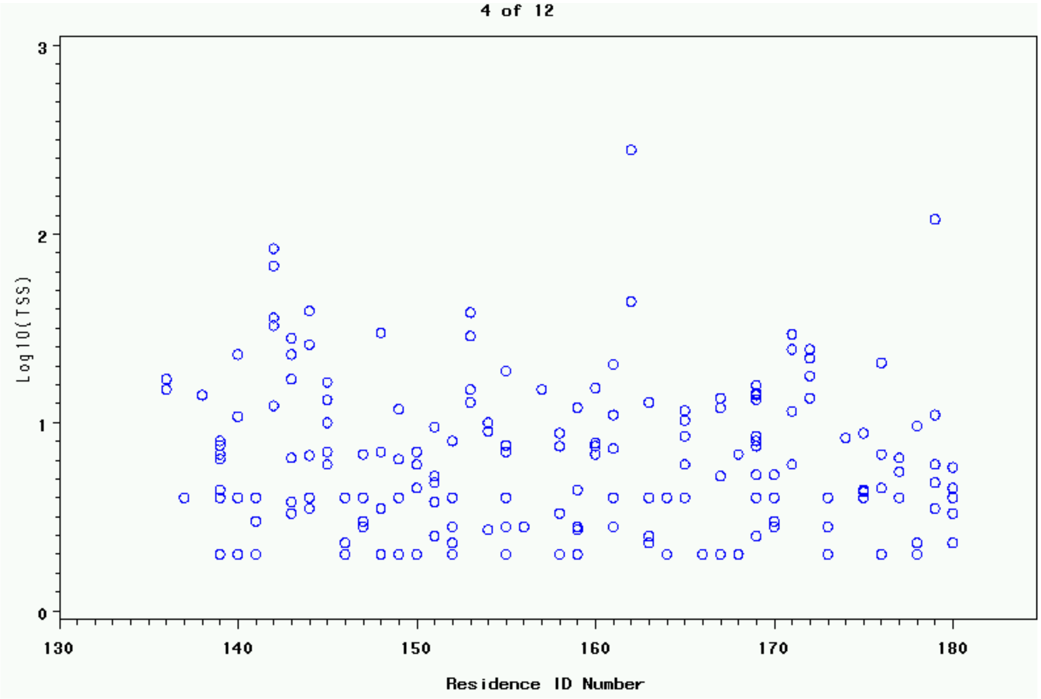


Figure 2.7.e: Distribution of FAST Log10(TSS) by Site

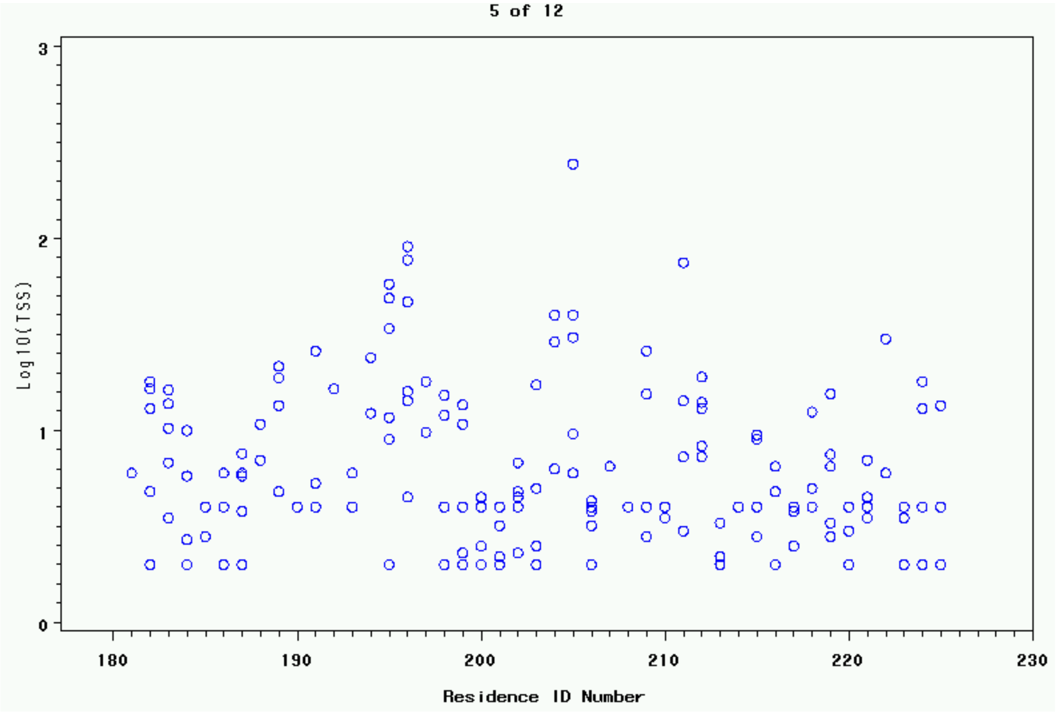


Figure 2.7.f: Distribution of FAST Log10(TSS) by Site

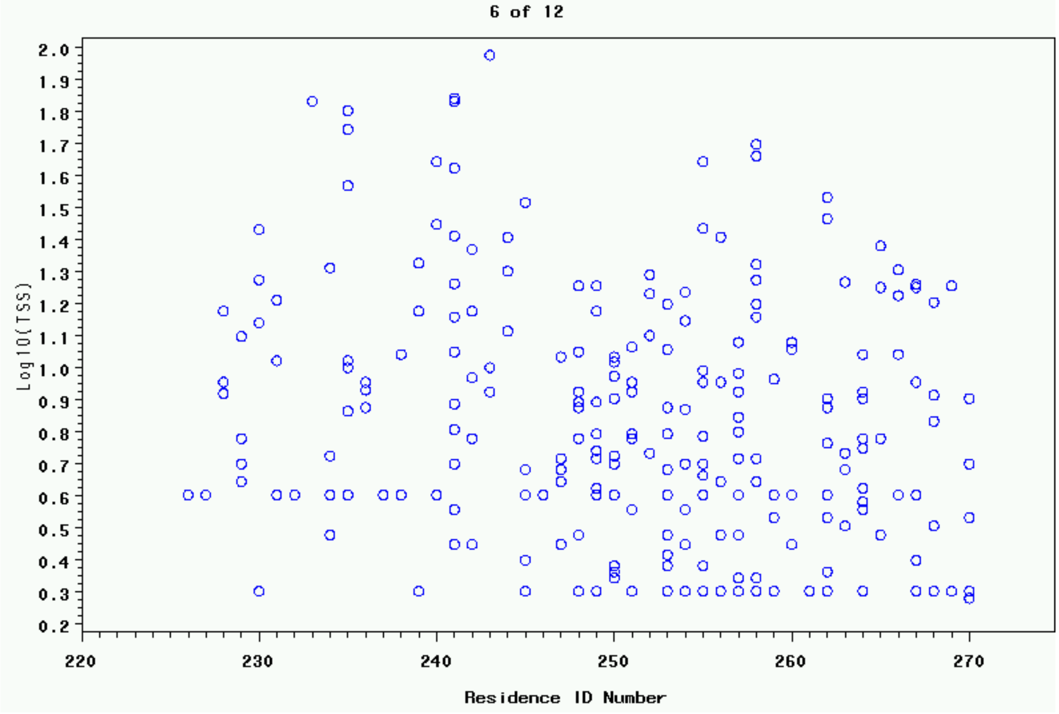


Figure 2.7.g: Distribution of FAST Log10(TSS) by Site

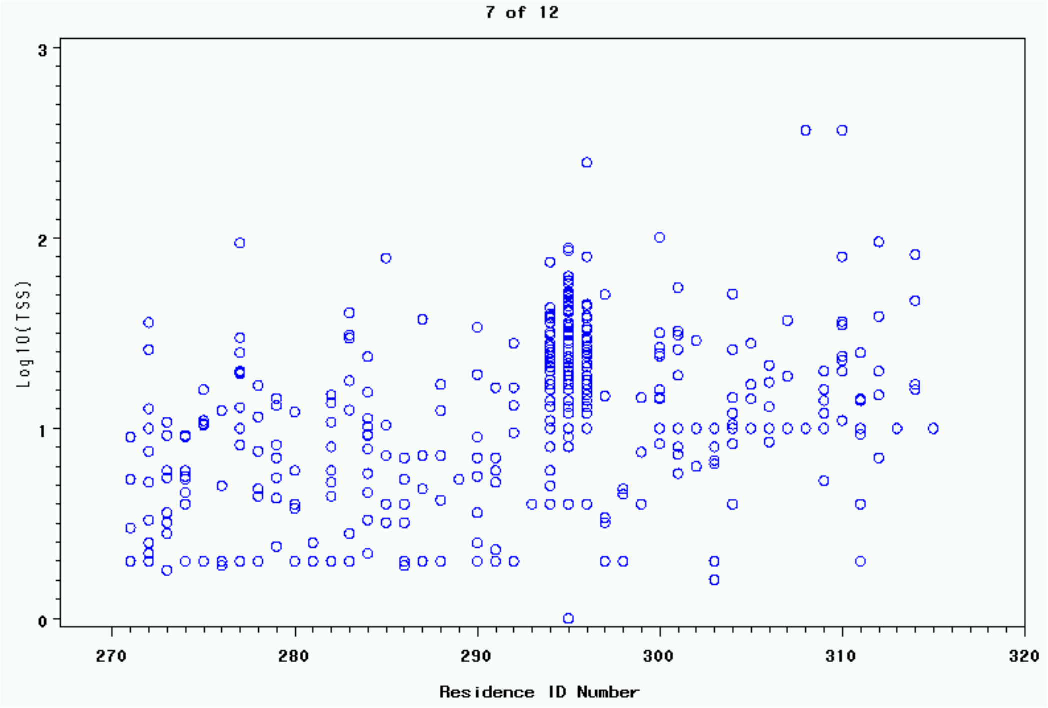


Figure 2.7.h: Distribution of FAST Log10(TSS) by Site

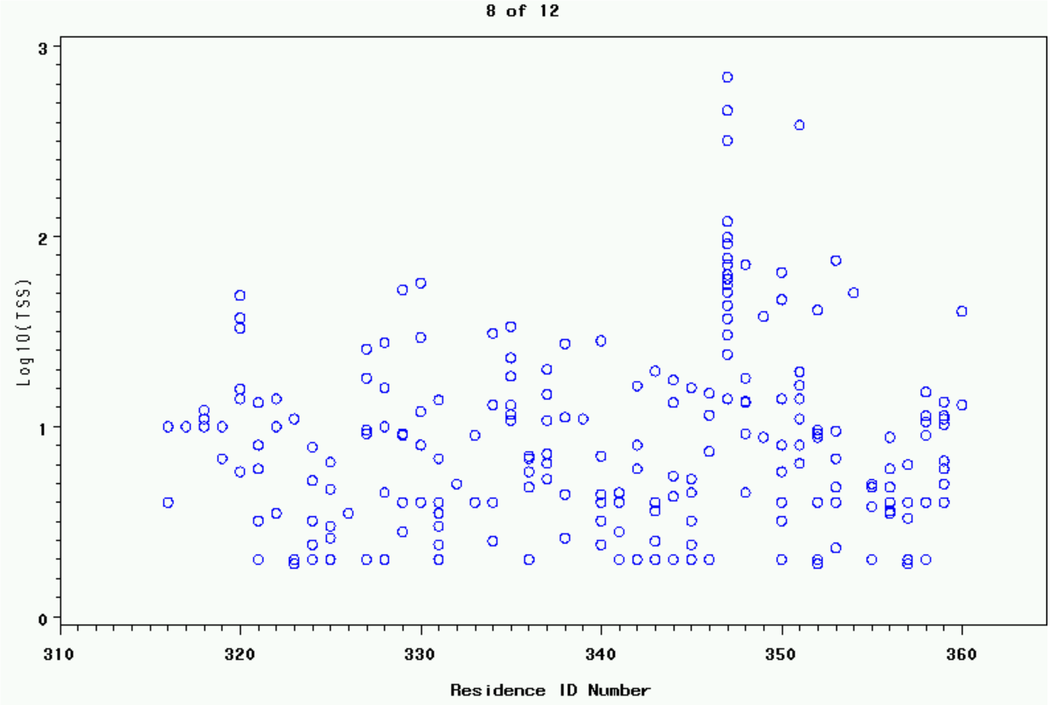


Figure 2.7.i: Distribution of FAST Log10(TSS) by Site

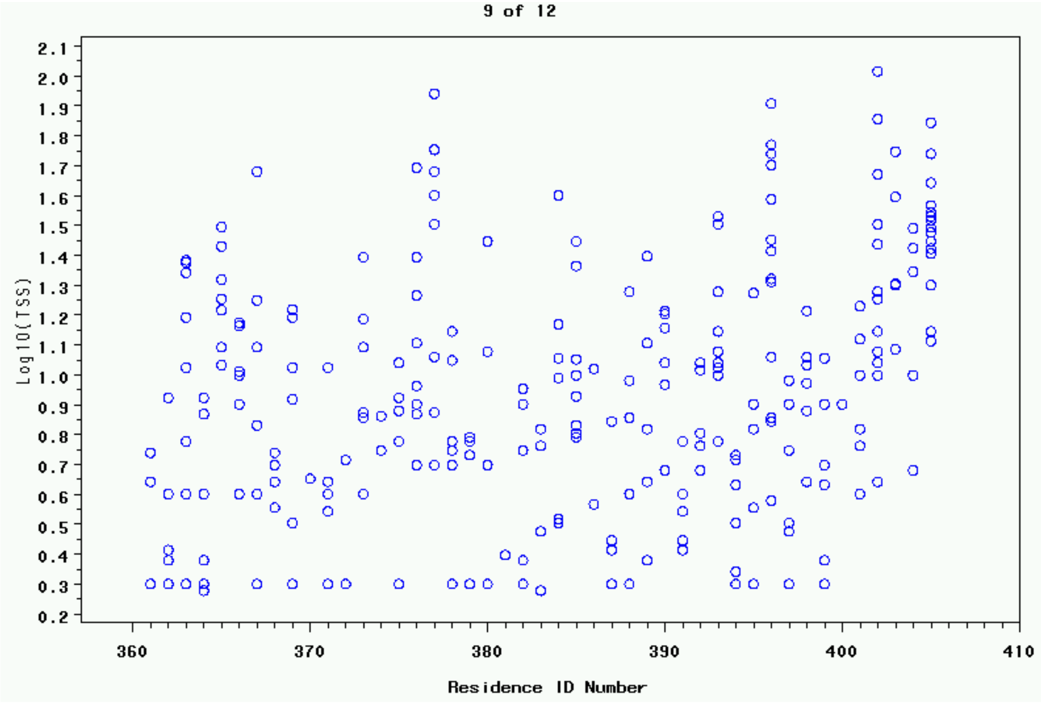


Figure 2.7.j: Distribution of FAST Log10(TSS) by Site

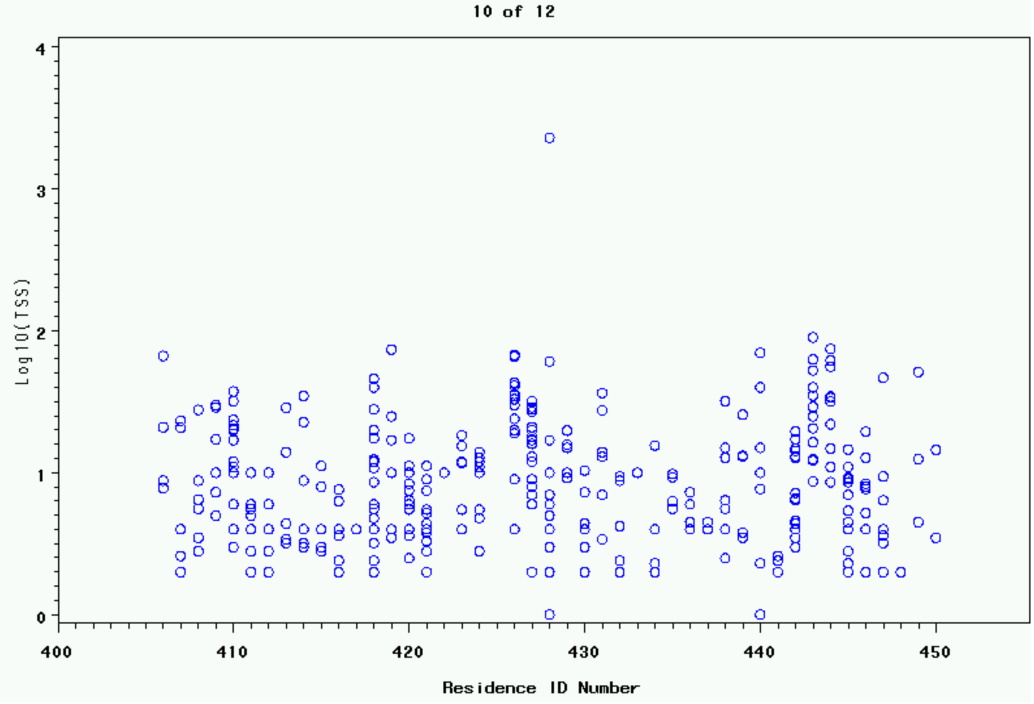


Figure 2.7.k: Distribution of FAST Log10(TSS) by Site

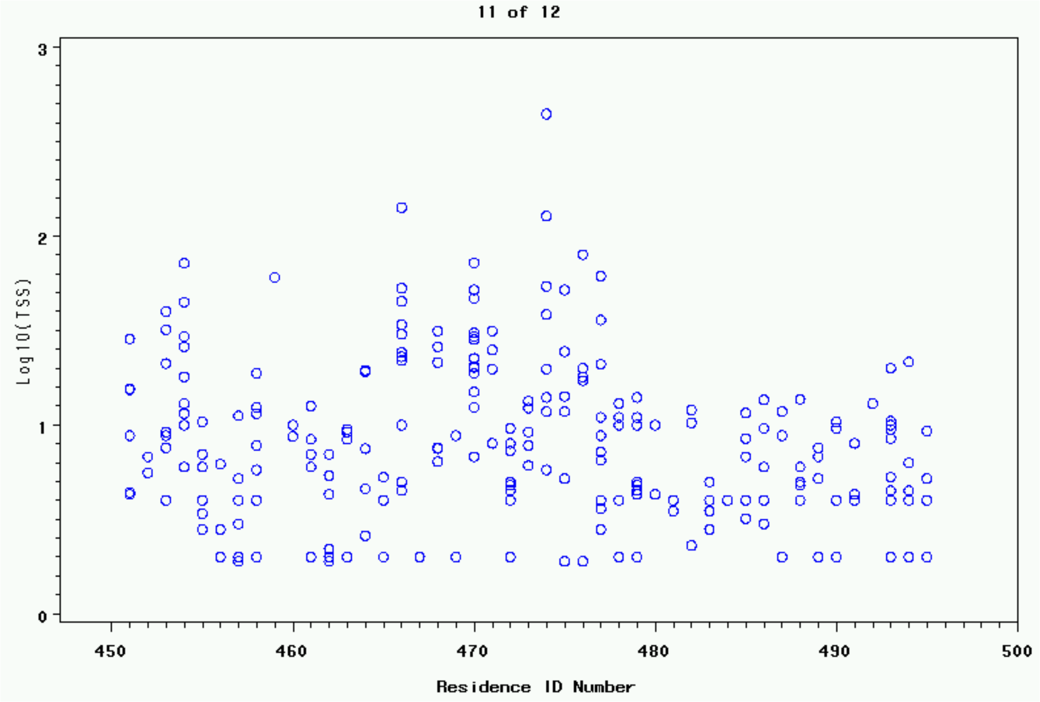


Figure 2.7.l: Distribution of FAST Log10(TSS) by Site

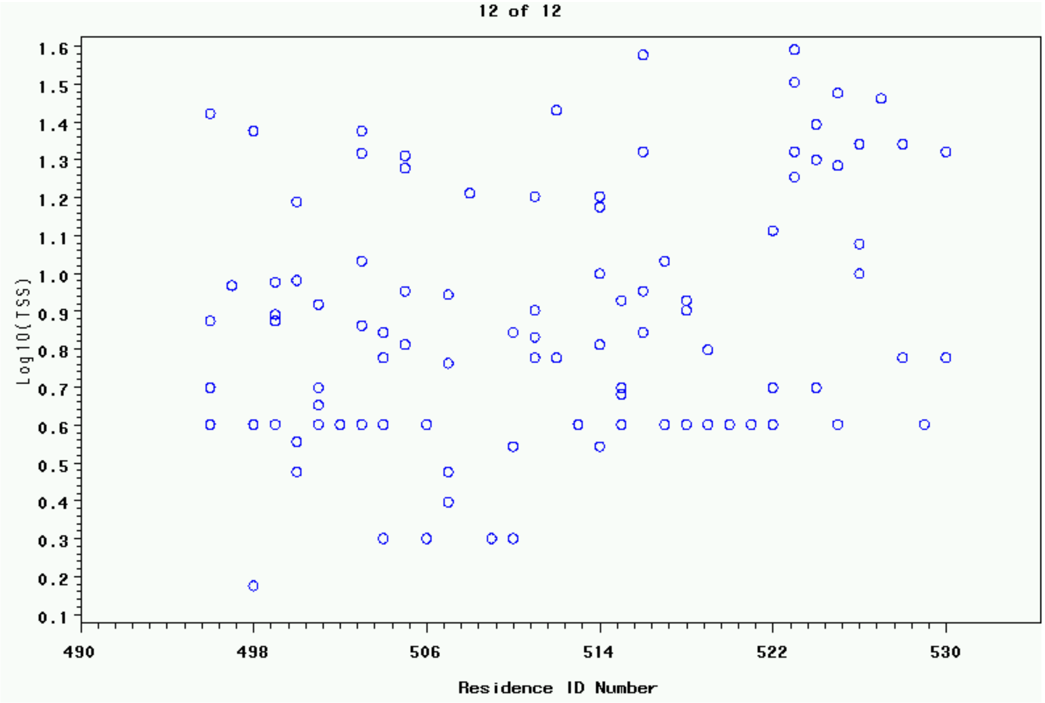


Figure 2.8.a: Distribution of FAST Log10(BOD) by Site

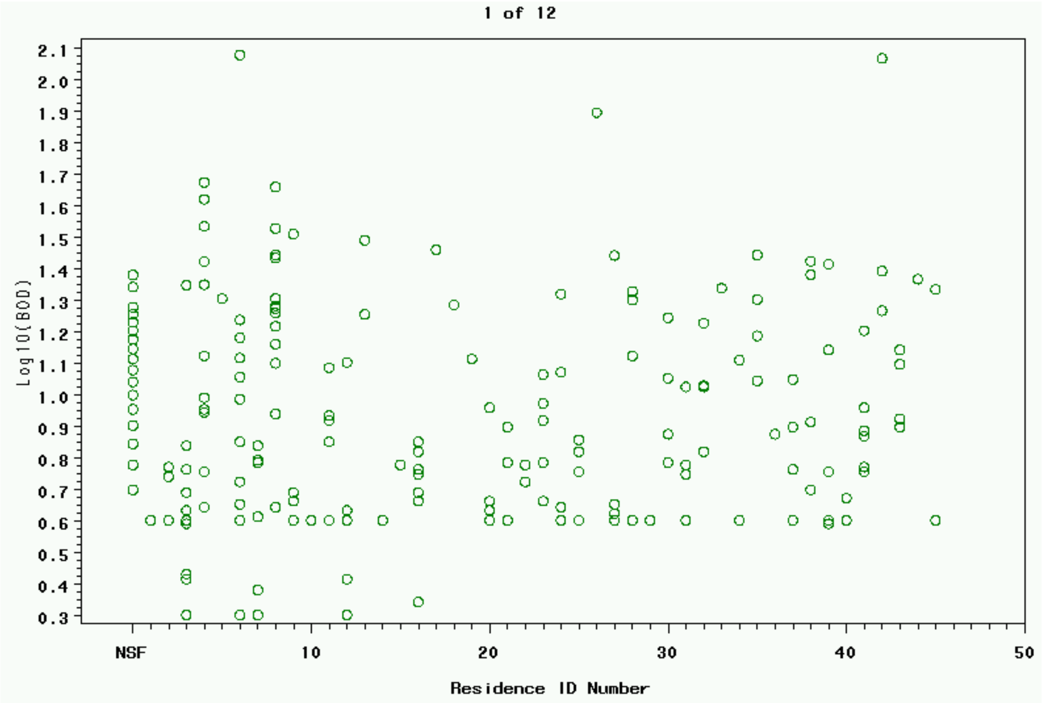


Figure 2.8.b: Distribution of FAST Log10(BOD) by Site

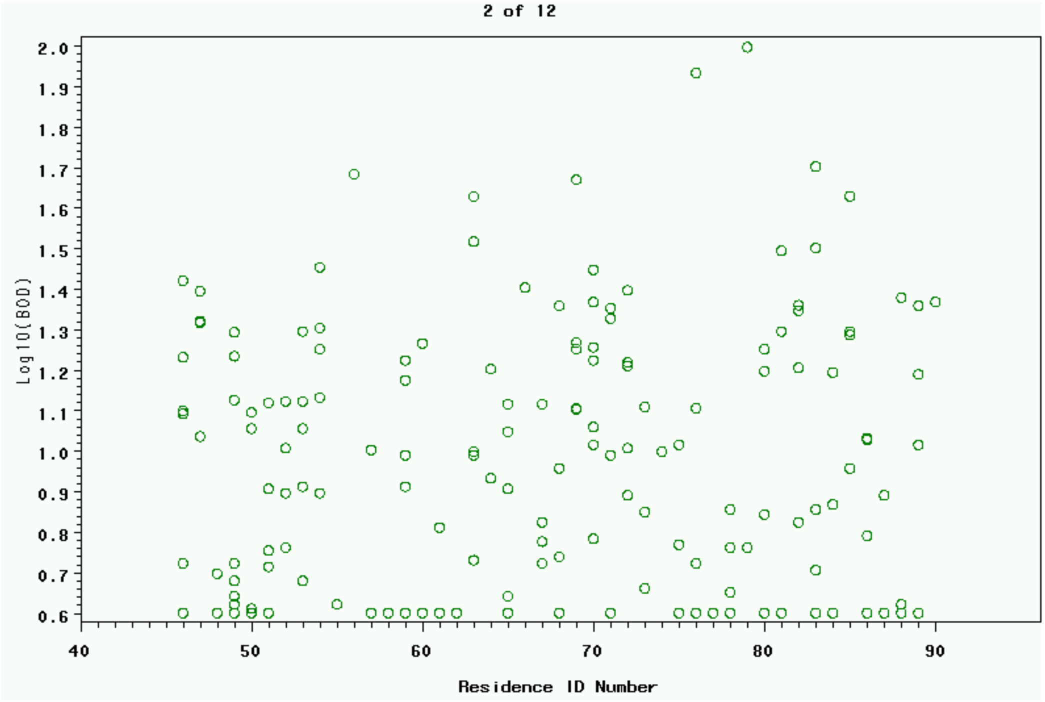


Figure 2.8.c: Distribution of FAST Log10(BOD) by Site

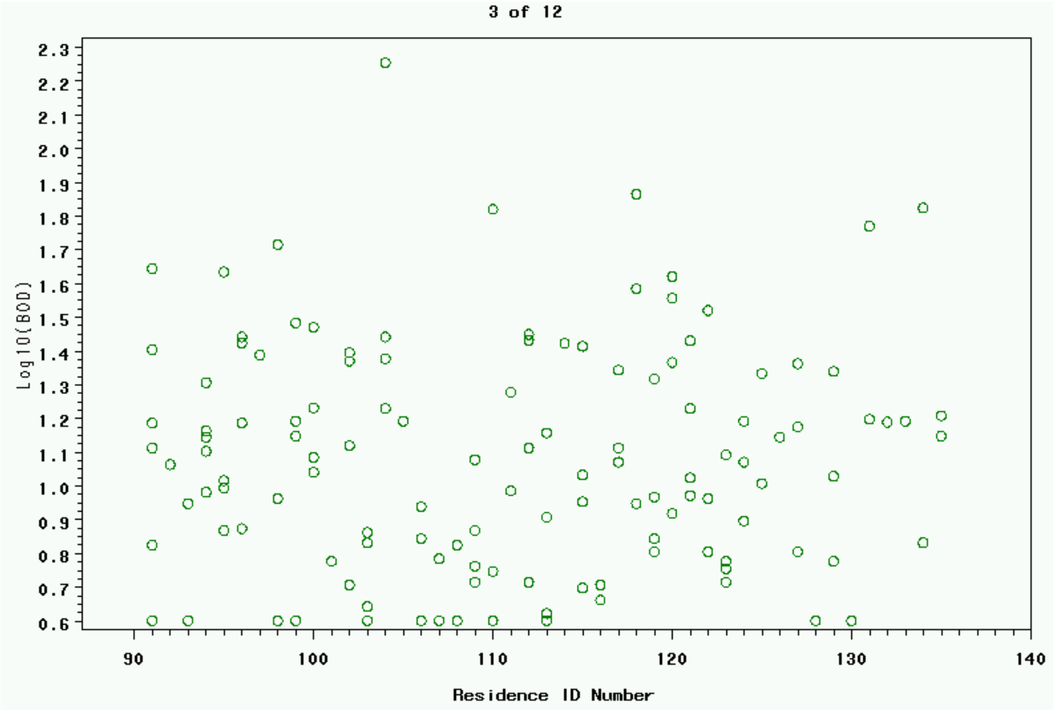


Figure 2.8.d: Distribution of FAST Log10(BOD) by Site

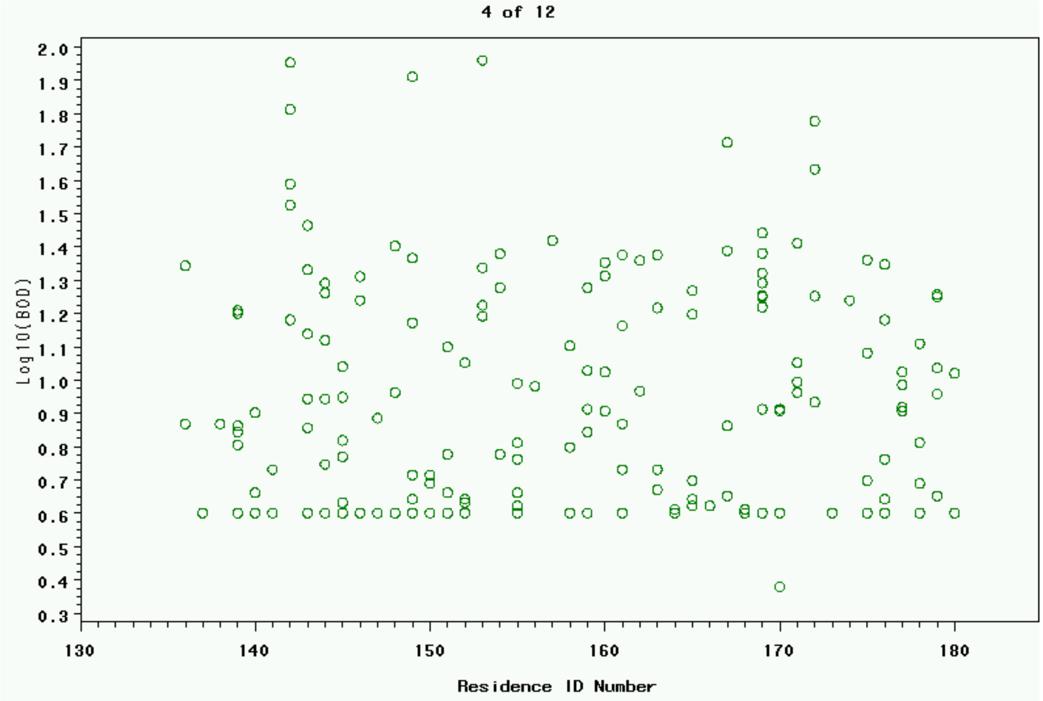


Figure 2.8.e: Distribution of FAST Log10(BOD) by Site

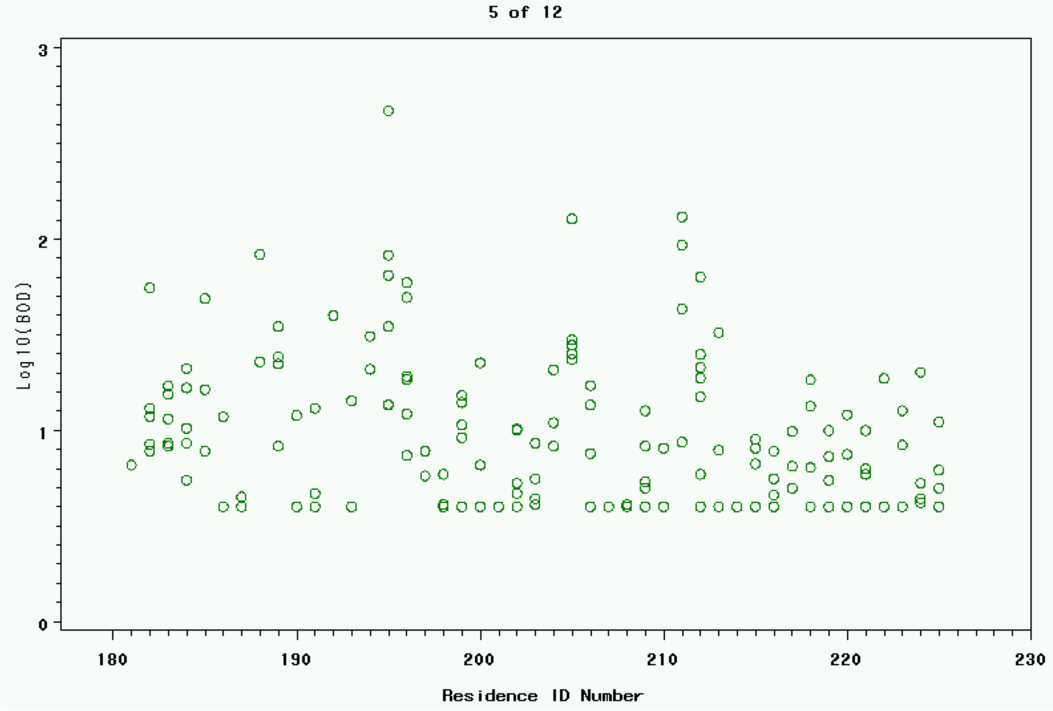


Figure 2.8.f: Distribution of FAST Log10(BOD) by Site

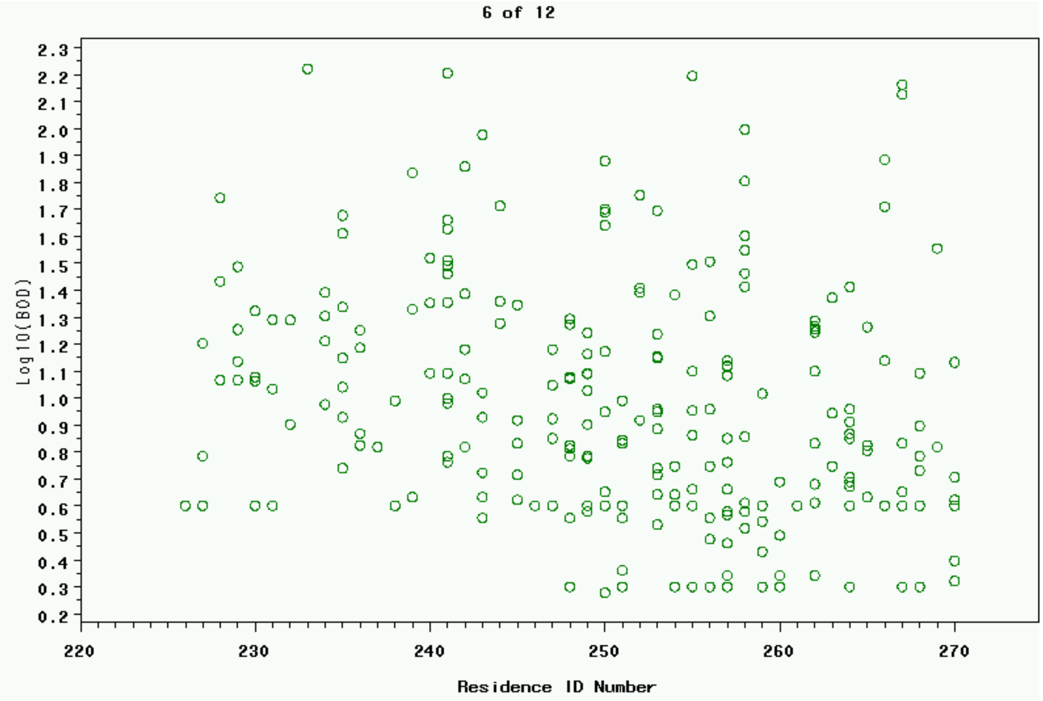


Figure 2.8.g: Distribution of FAST Log10(BOD) by Site

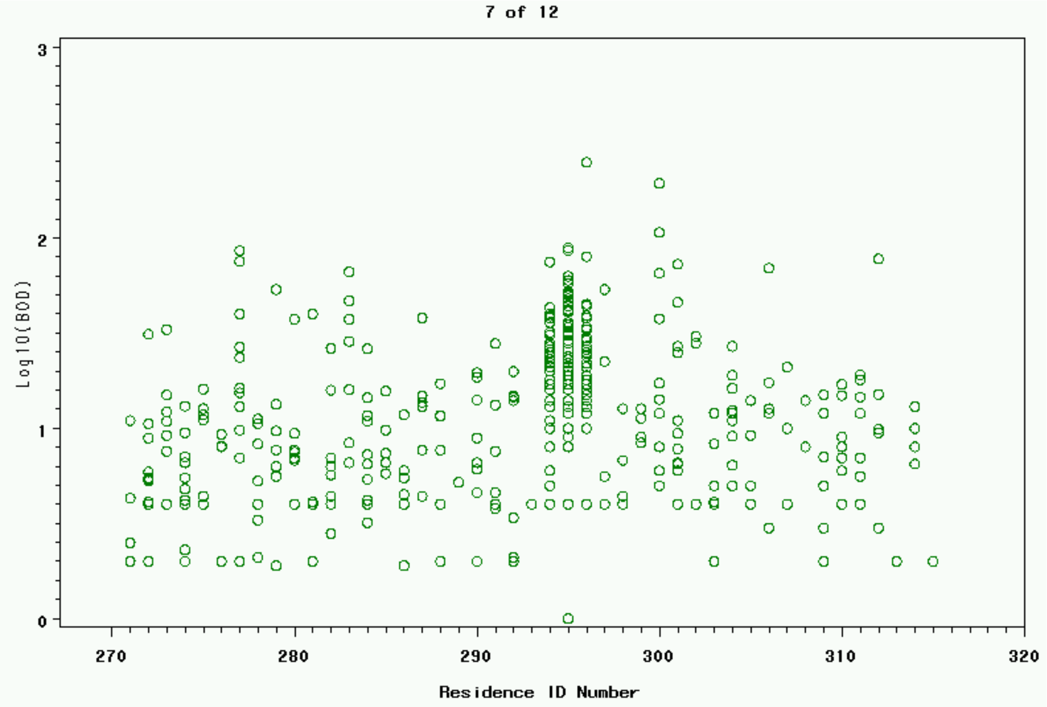


Figure 2.8.h: Distribution of FAST Log10(BOD) by Site

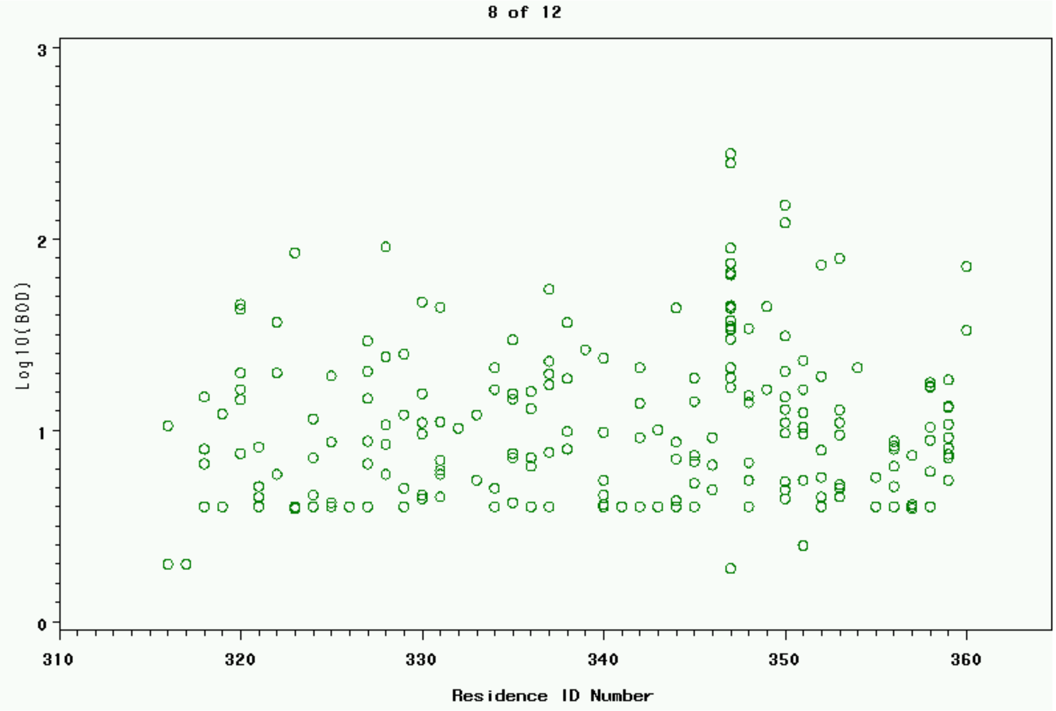


Figure 2.8.i: Distribution of FAST Log10(BOD) by Site

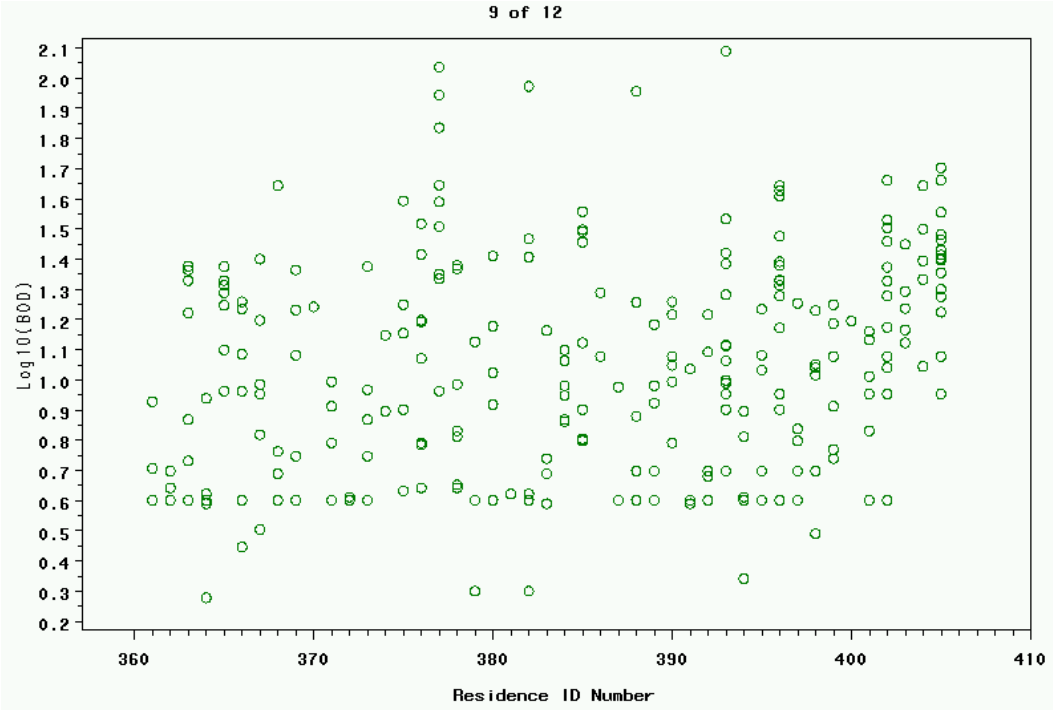


Figure 2.8.j: Distribution of FAST Log10(BOD) by Site

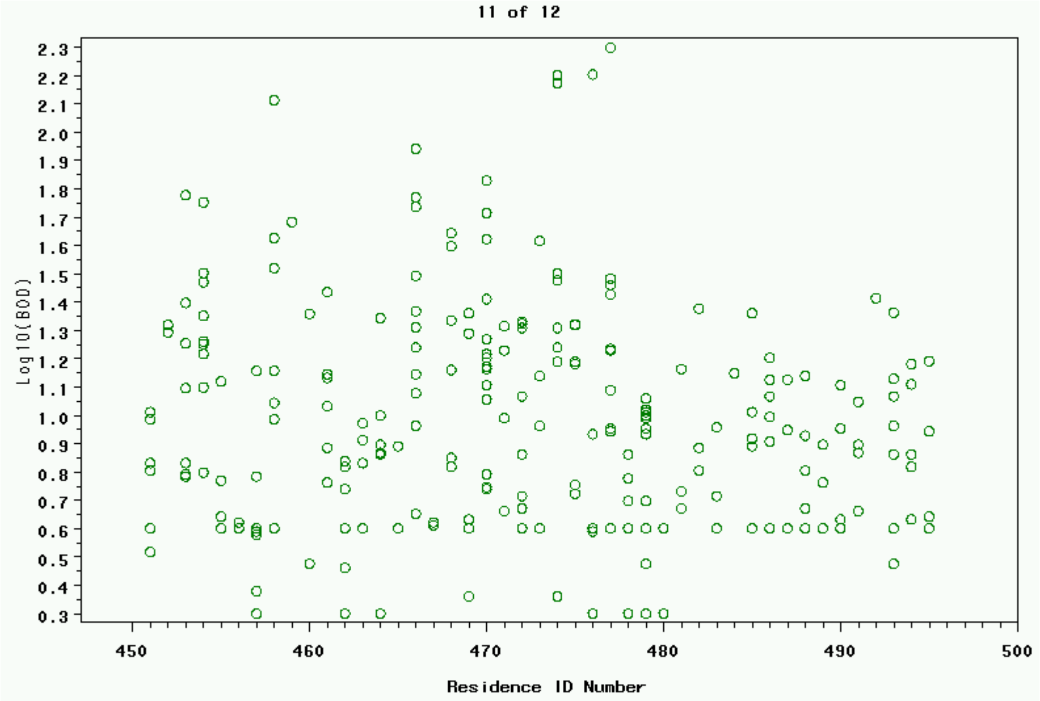


Figure 2.8.k: Distribution of FAST Log10(BOD) by Site

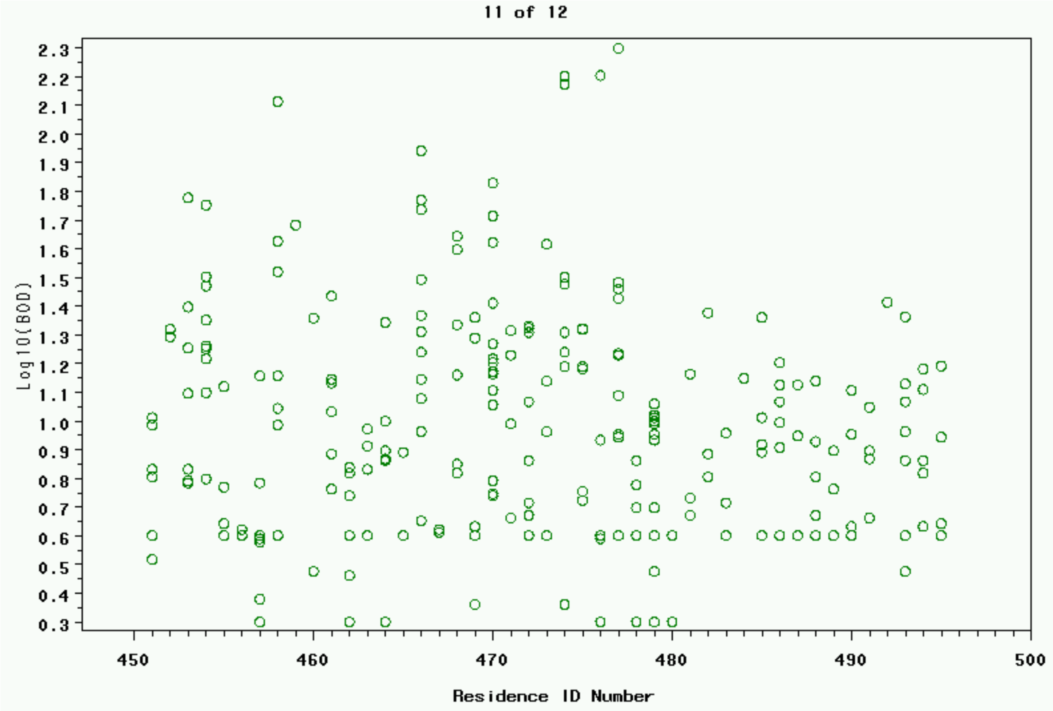


Figure 2.8.l: Distribution of FAST Log10(BOD) by Site

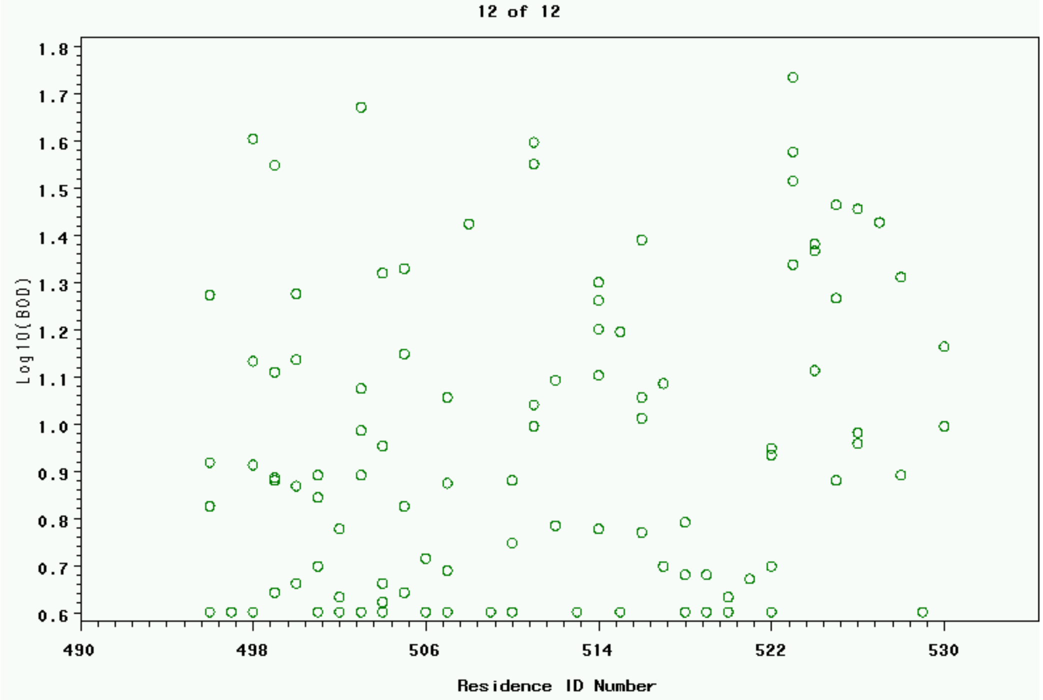


Figure 2.9.a. Frequency Distribution of Number of Observations at a Site: Advantex Technology TSS.

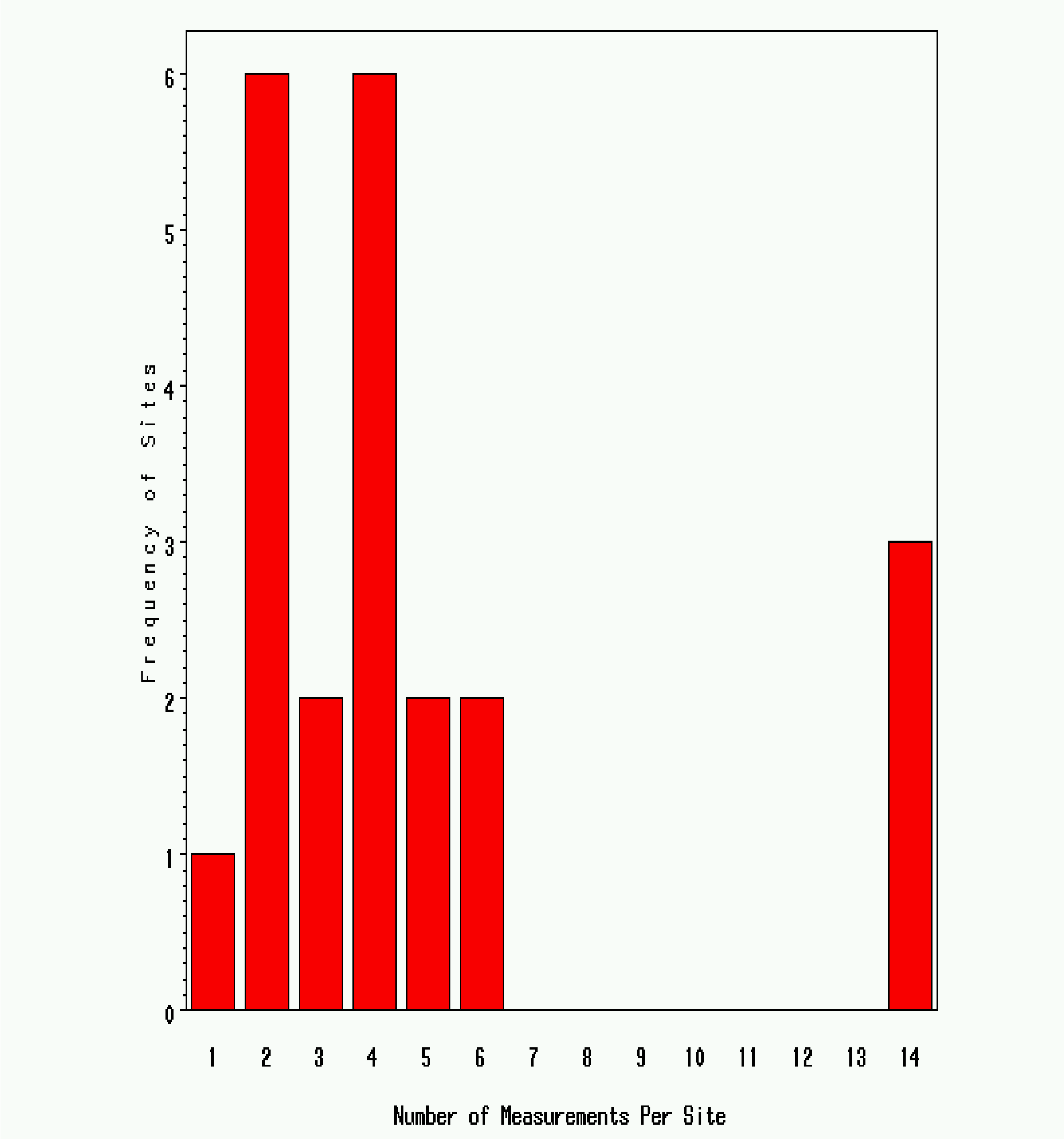


Figure 2.9.b. Frequency Distribution of Number of Observations at a Site: Advantex Technology BOD.

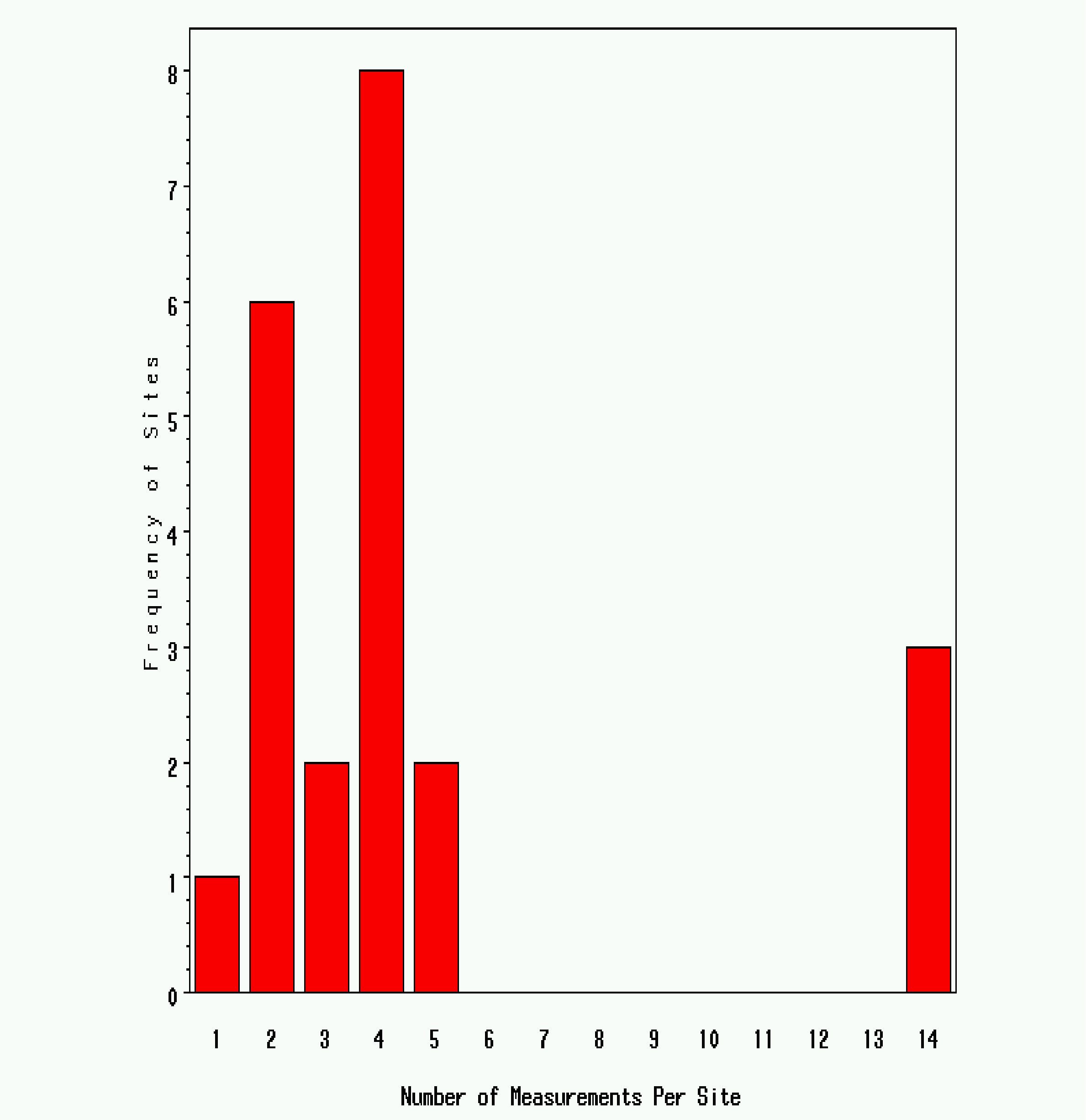


Figure 2.10.a. Frequency Distribution of Number of Observations at a Site: Bioclere Technology TSS.

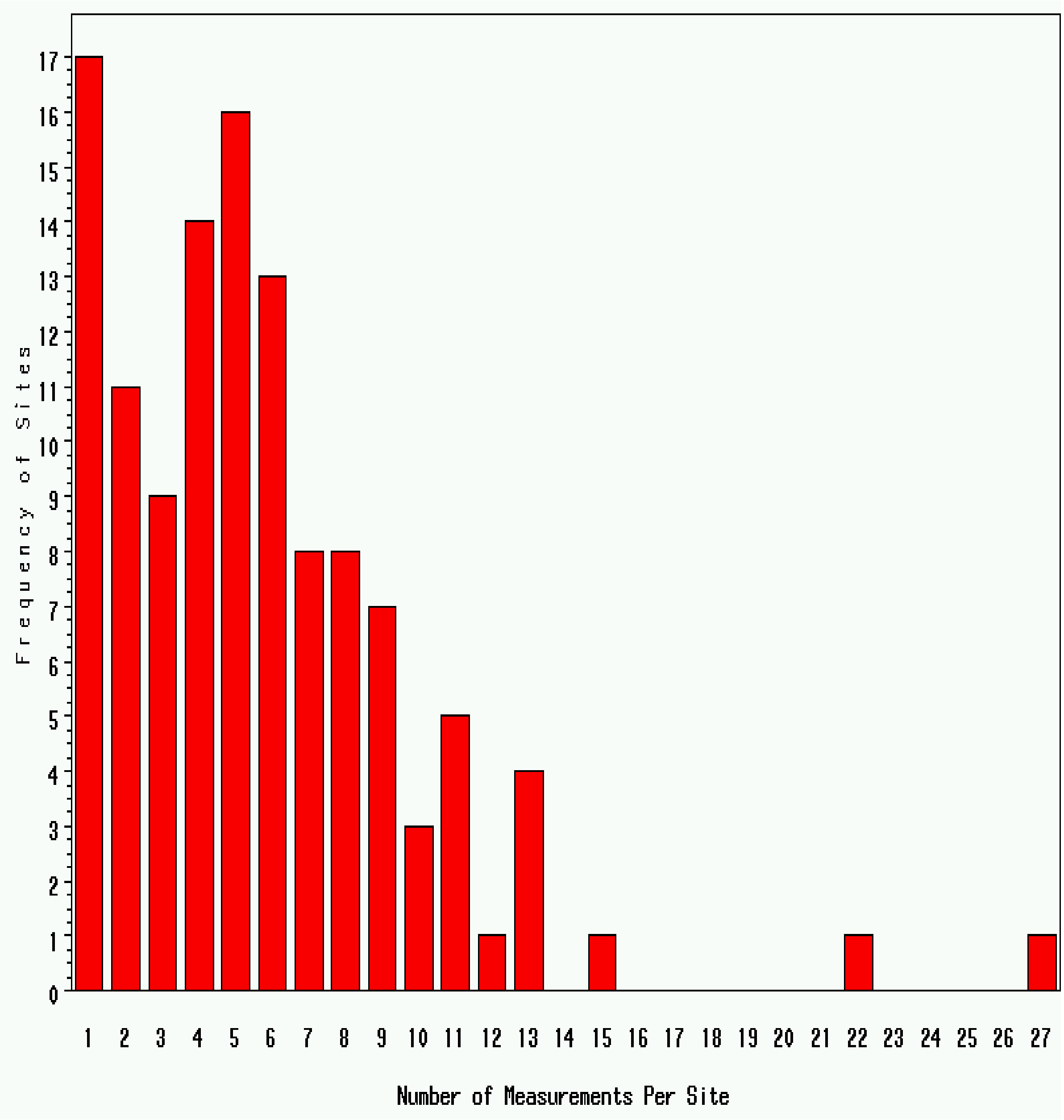


Figure 2.10.b. Frequency Distribution of Number of Observations at a Site: Bioclere Technology BOD.

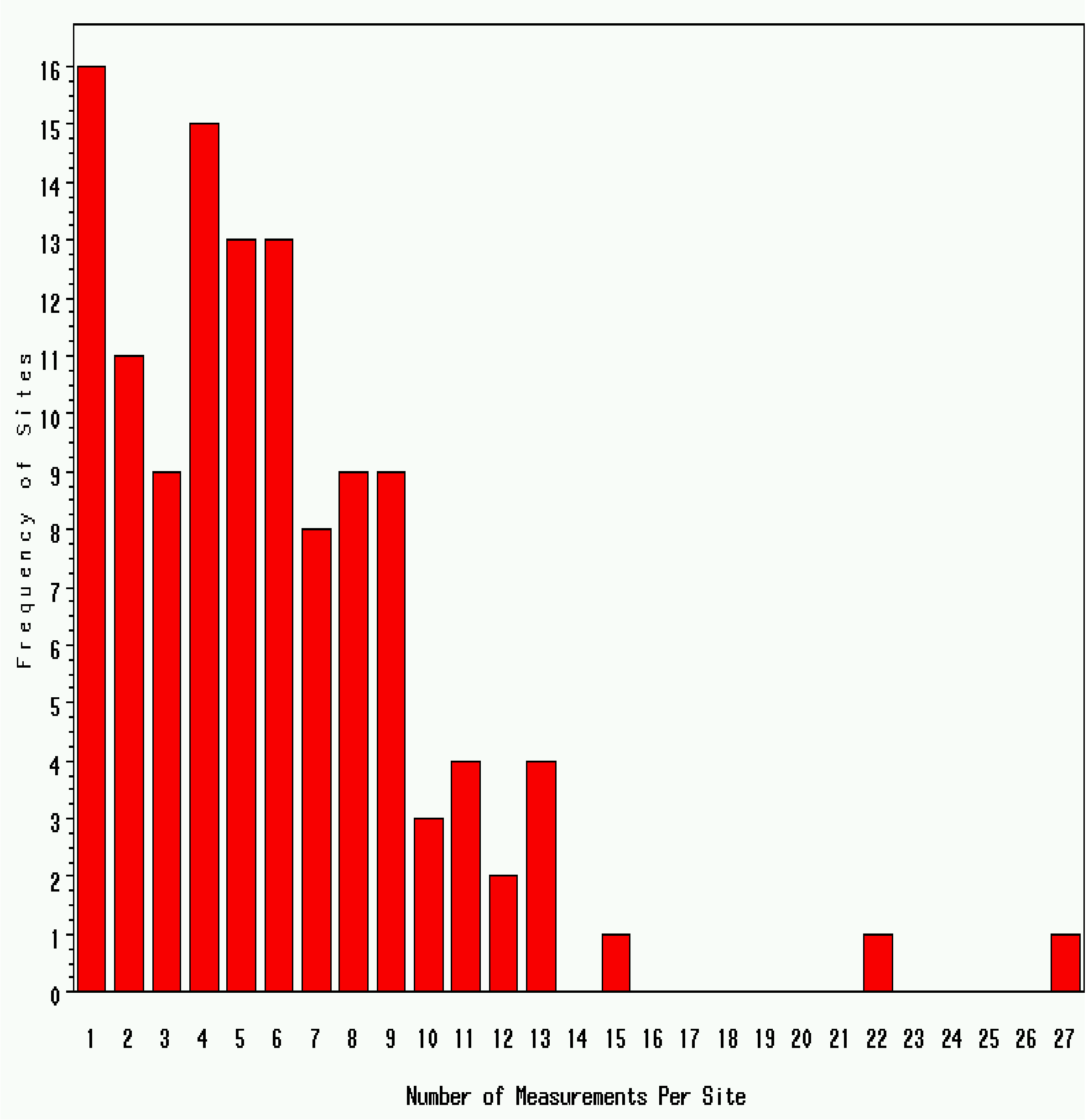


Figure 2.11.a. Frequency Distribution of Number of Observations at a Site: FAST Technology TSS.

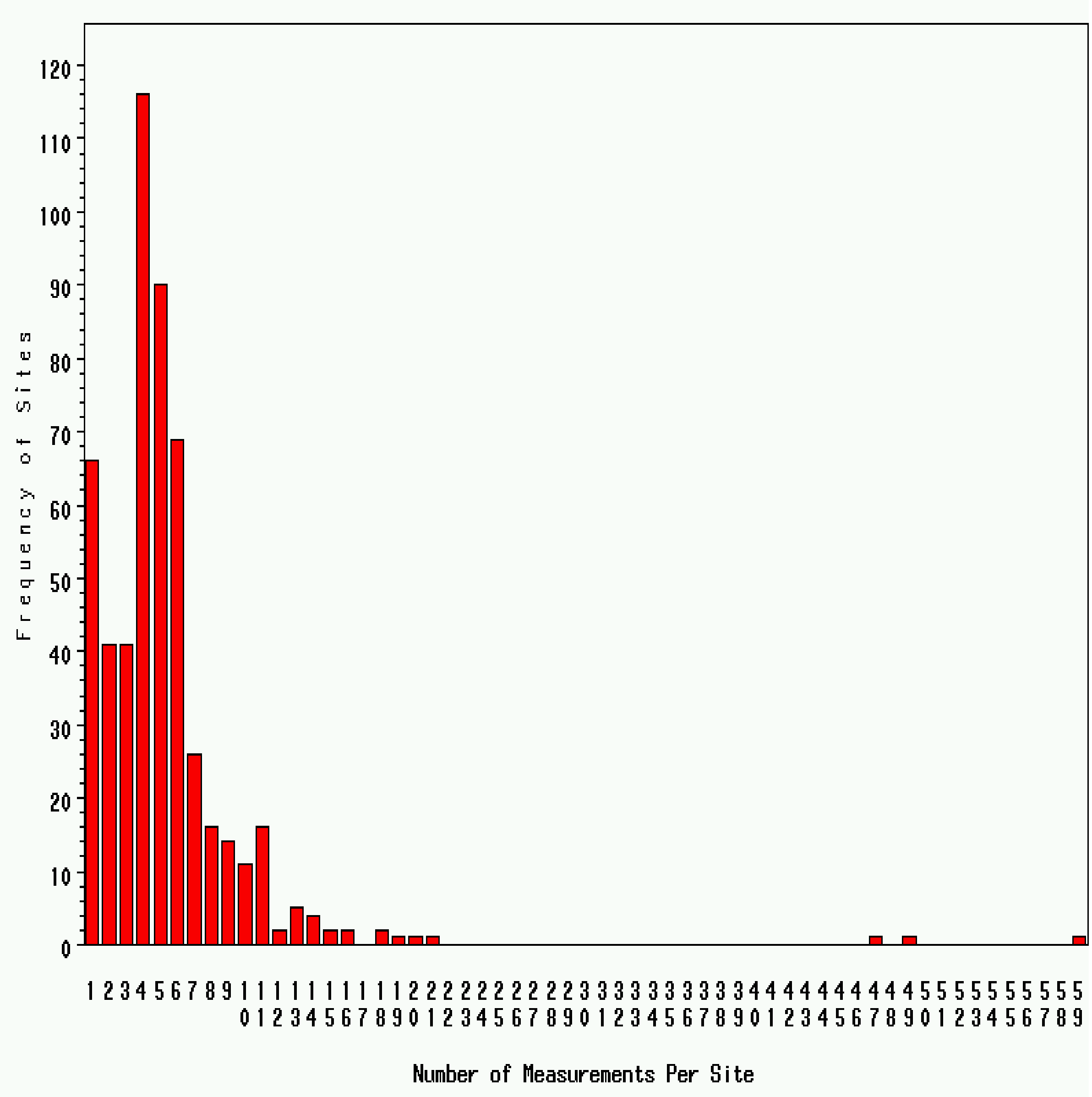


Figure 2.11.b. Frequency Distribution of Number of Observations at a Site: FAST Technology TSS.

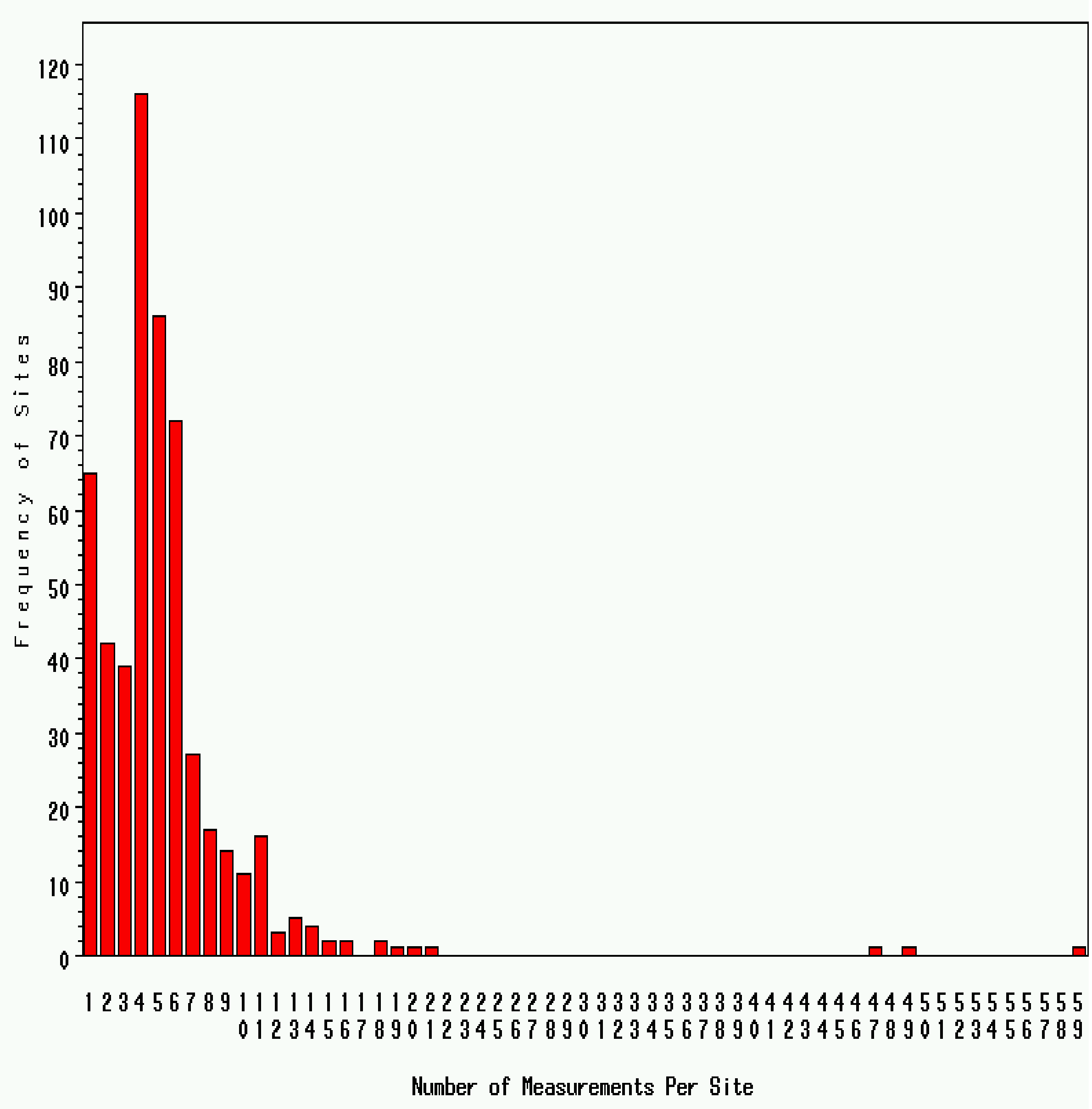


Figure 2.12.a. BOD Versus Time for Advantex Technology.

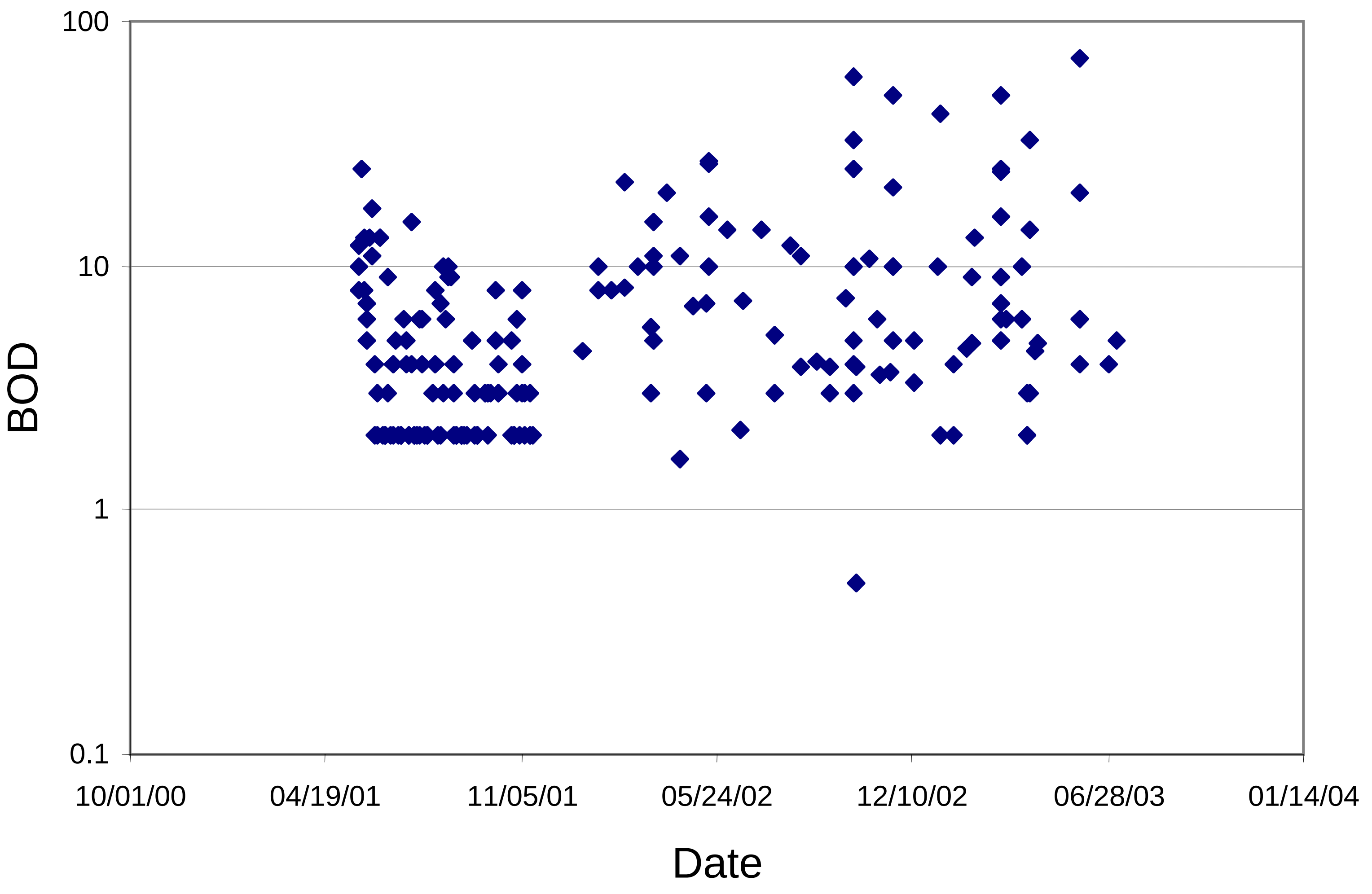


Figure 2.12.b. TSS Versus Time for Advantex Technology

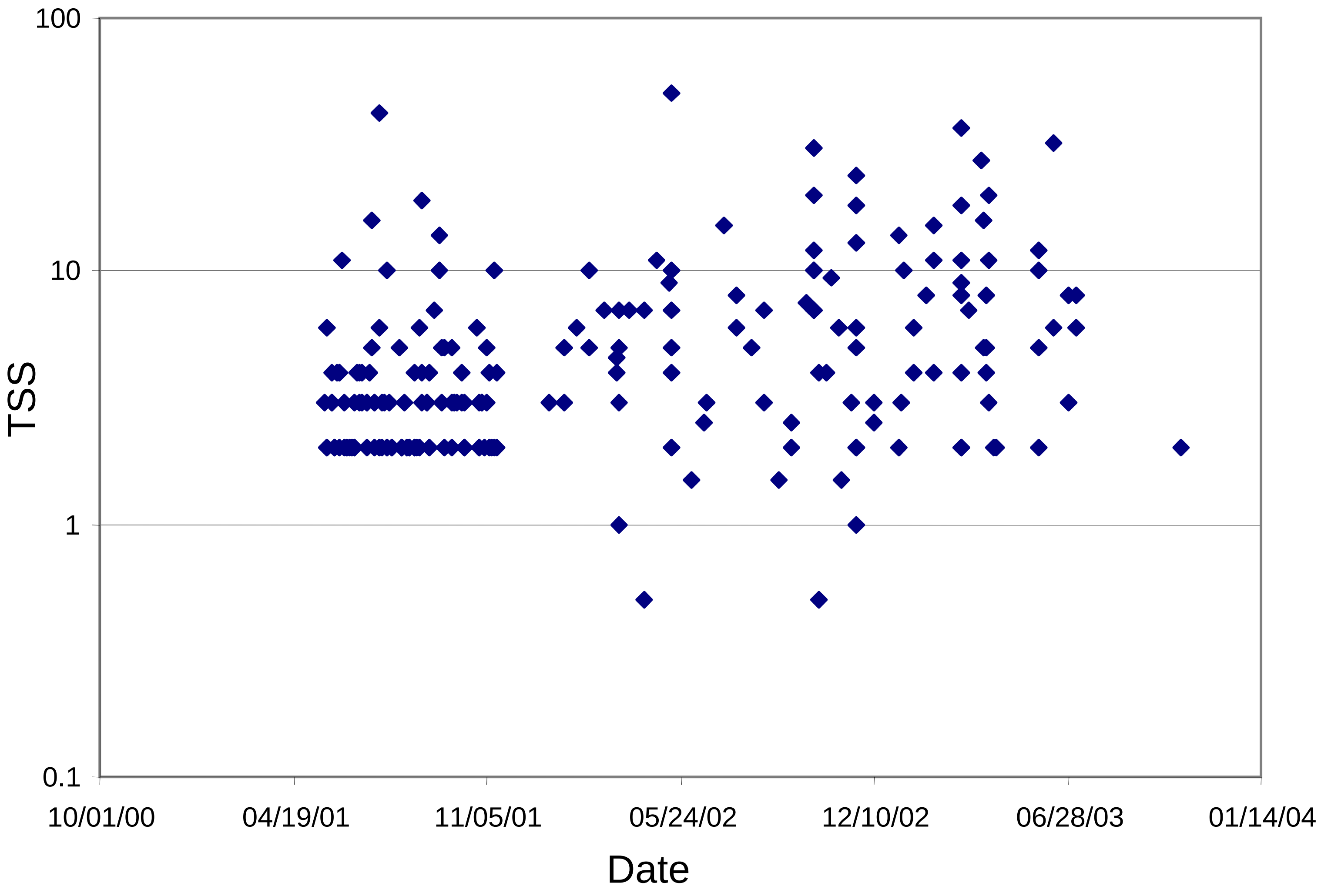


Figure 2.13.a. BOD Versus Time for Bioclere Technology.

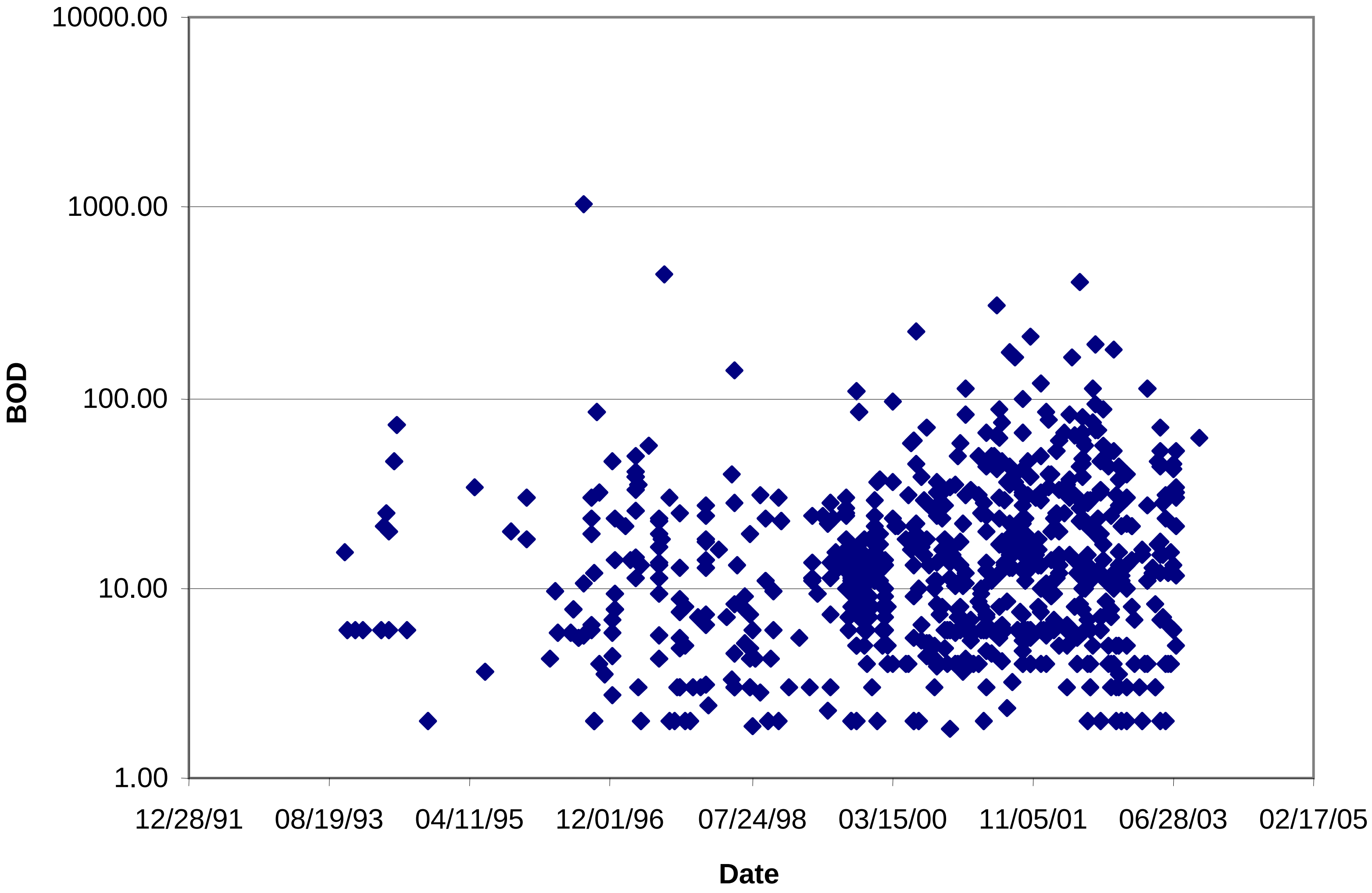


Figure 2.13.b. TSS Versus Time for Bioclere Technology.

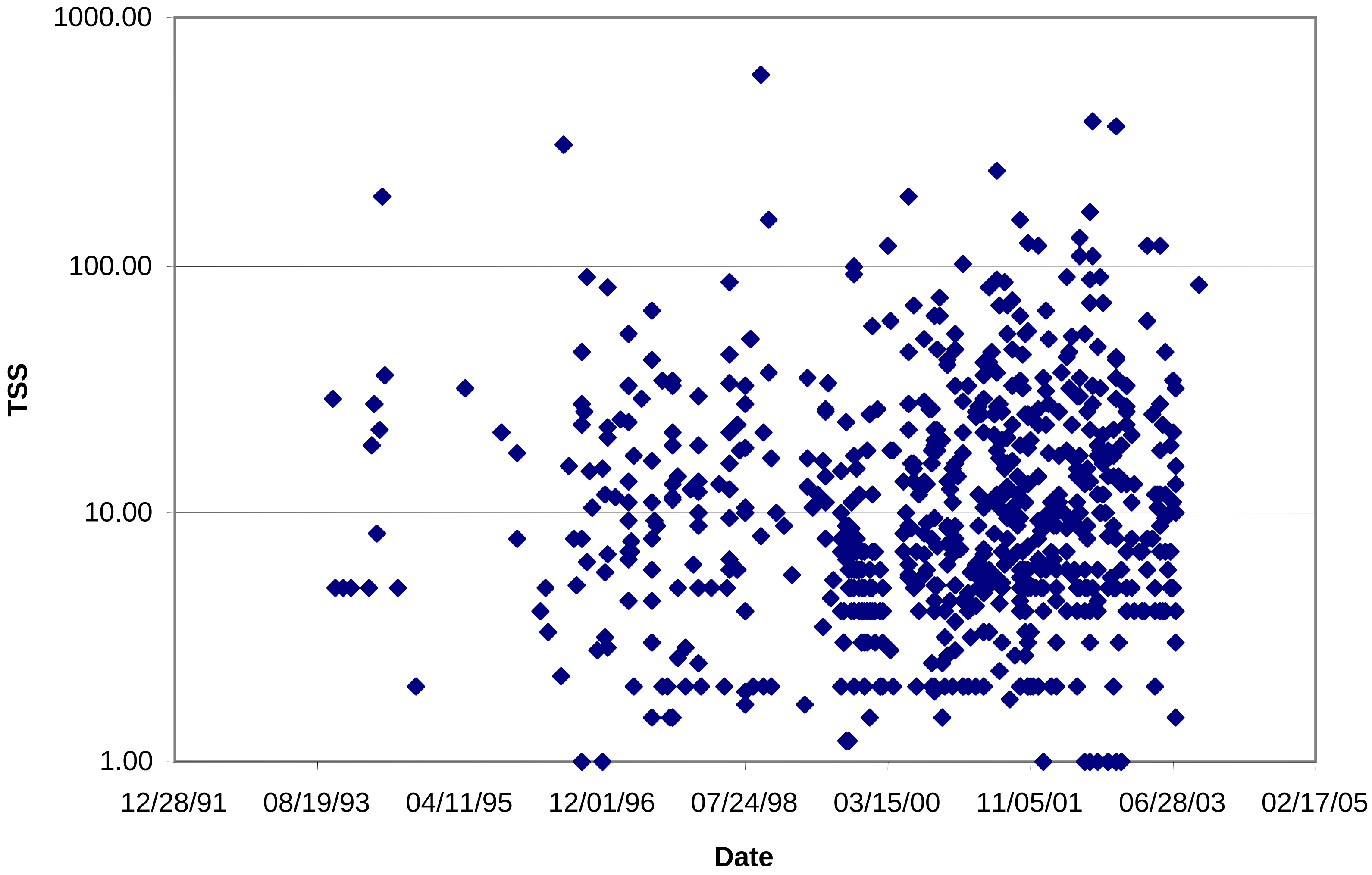


Figure 2.14.a. BOD Versus Time for FAST Technology.

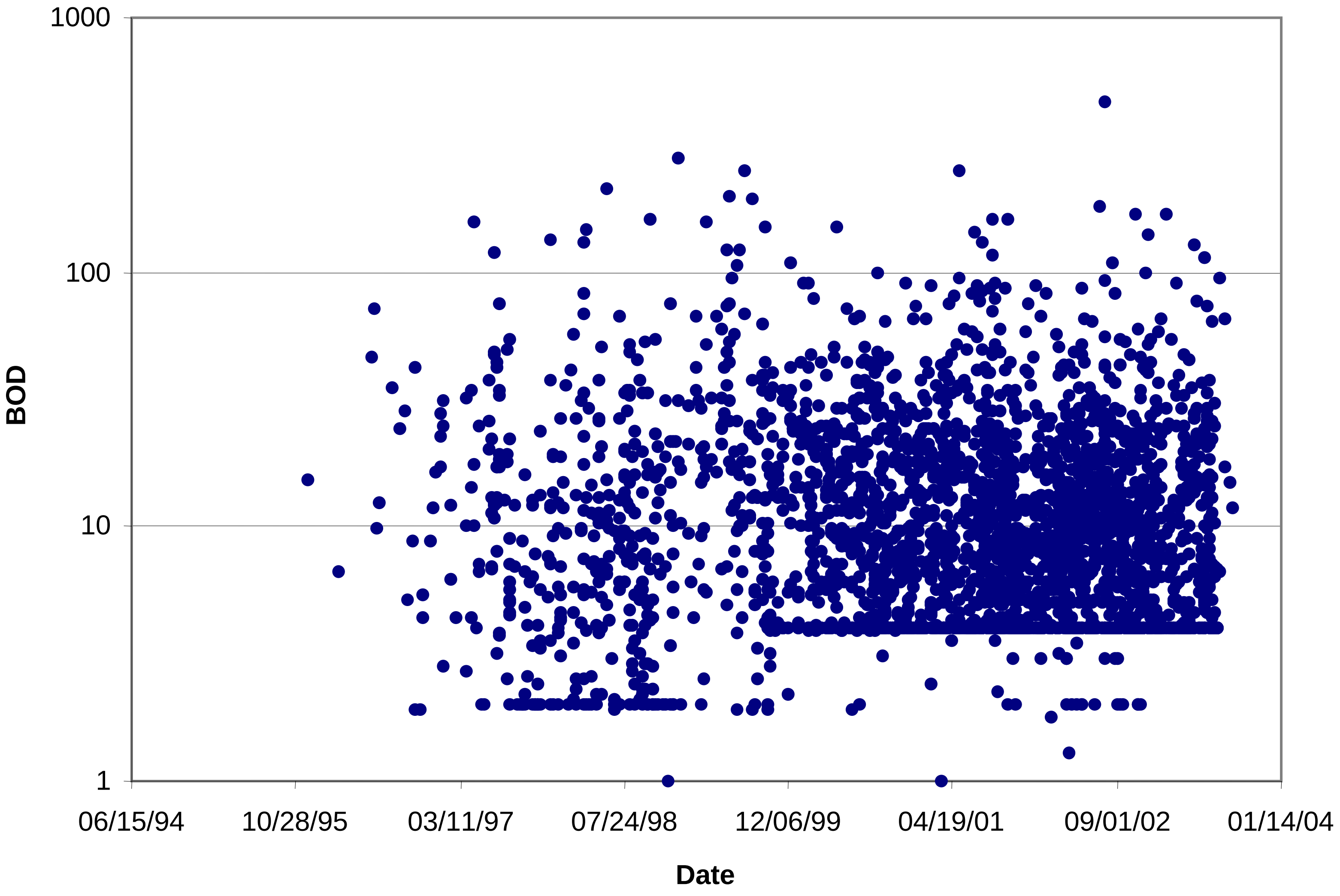


Figure 2.14.b. TSS Versus Time for FAST Technology

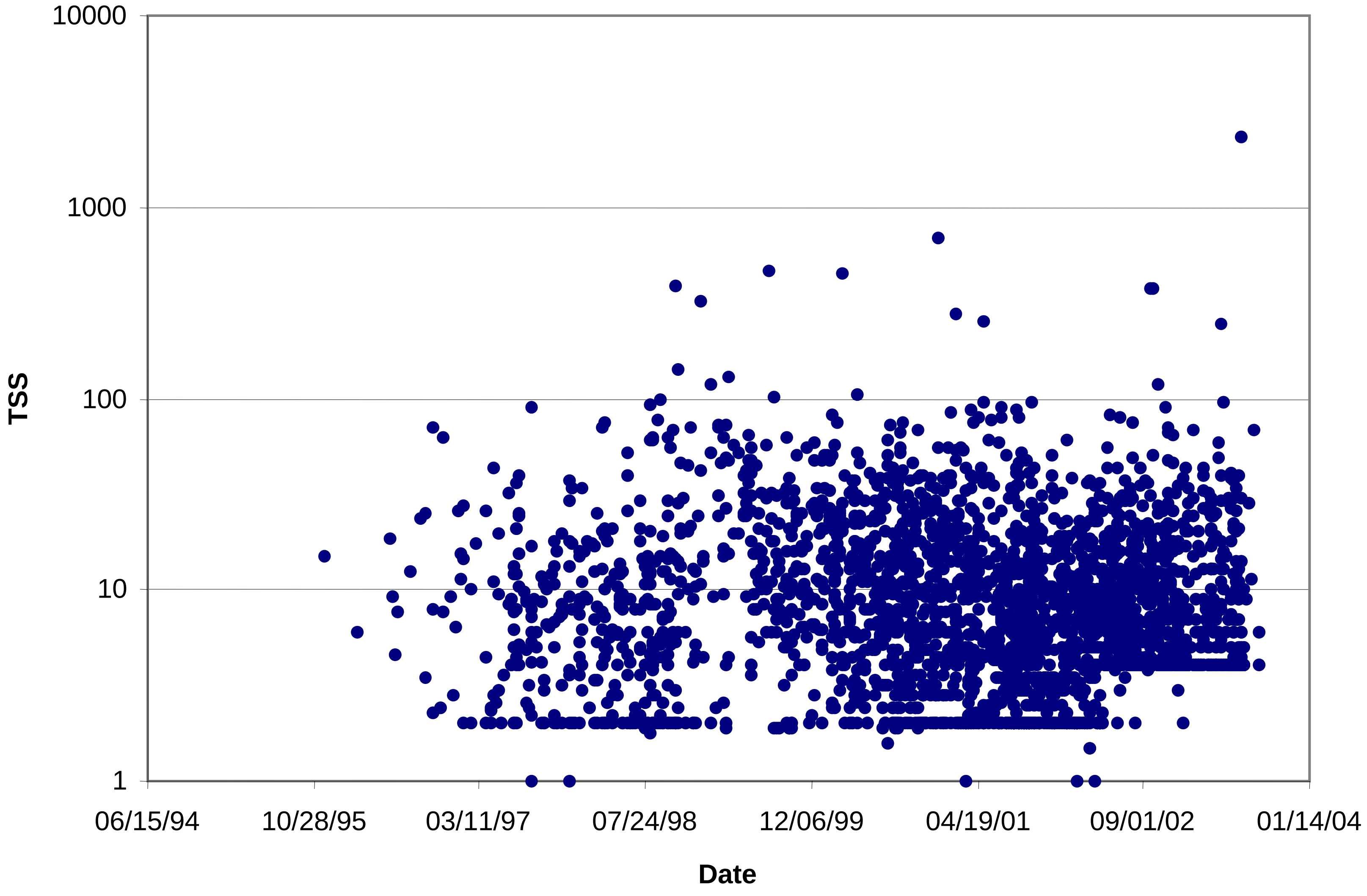


Figure 2.15.a. TSS Versus BOD for Advantex Technology.

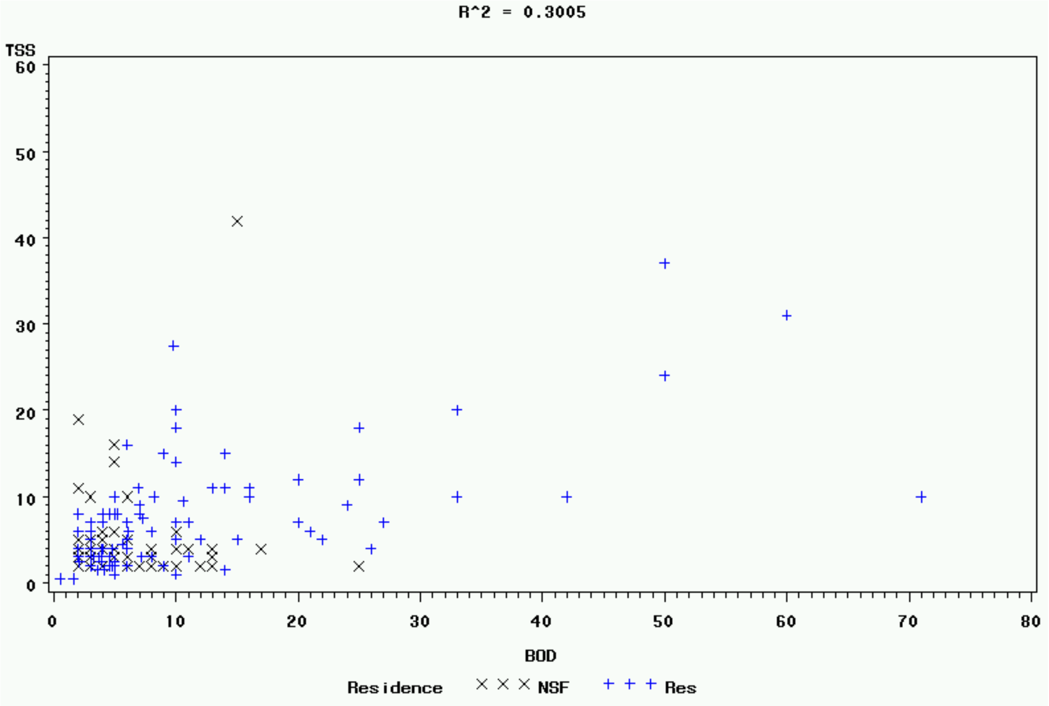


Figure 2.15.b. TSS Versus BOD for Advantex Technology (Log10)

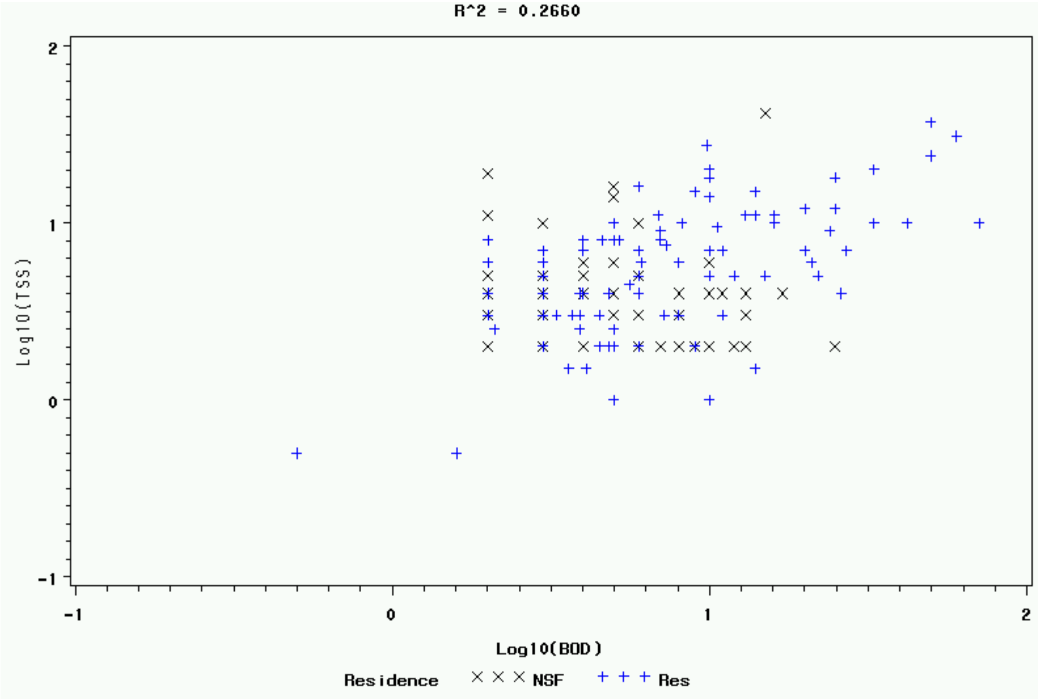


Figure 2.16.a. TSS Versus BOD for Bioclere Technology.

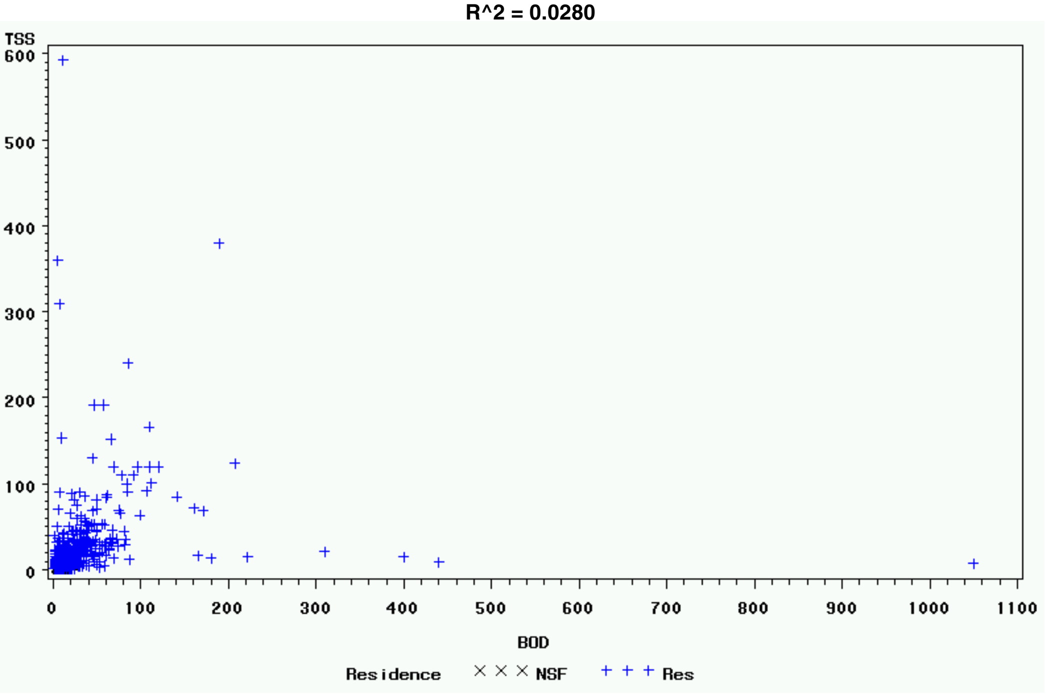


Figure 2.16.b. TSS Versus BOD for Bioclere (Log10).

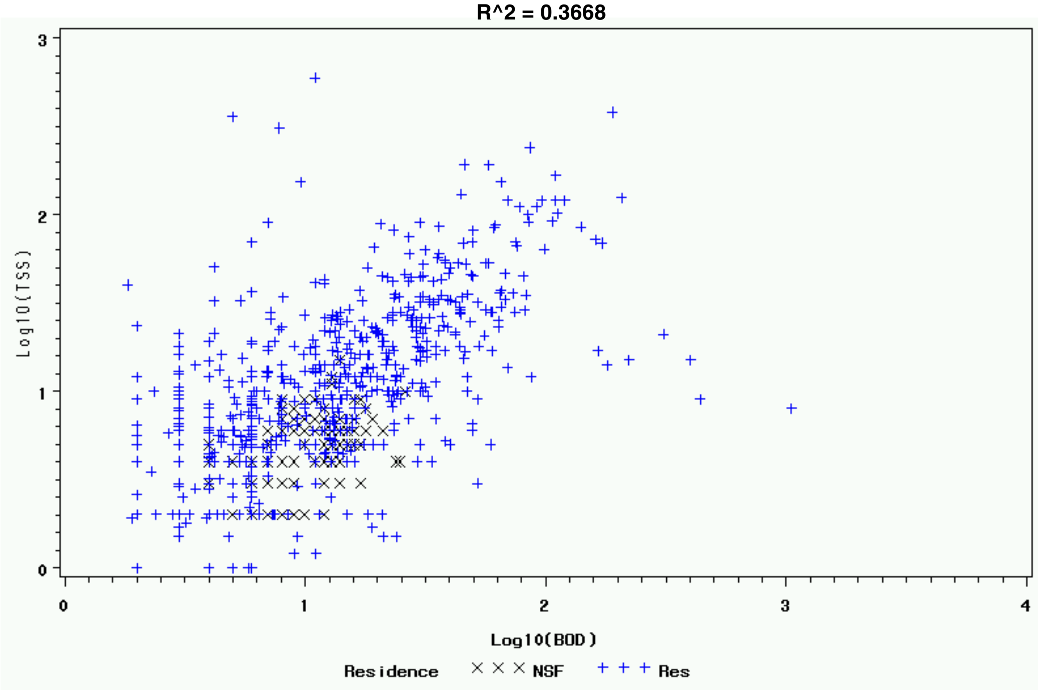


Figure 2.17. TSS Versus BOD for FAST Technology.

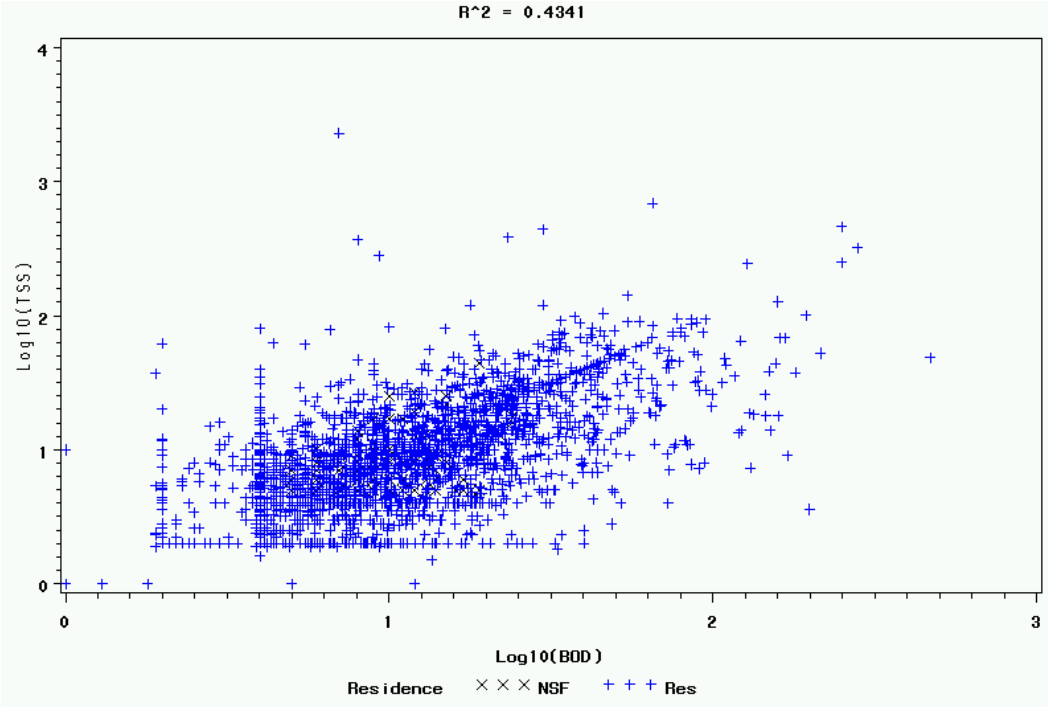


Figure 2.18.a. Monthly Means for TSS: Advantex Technology.

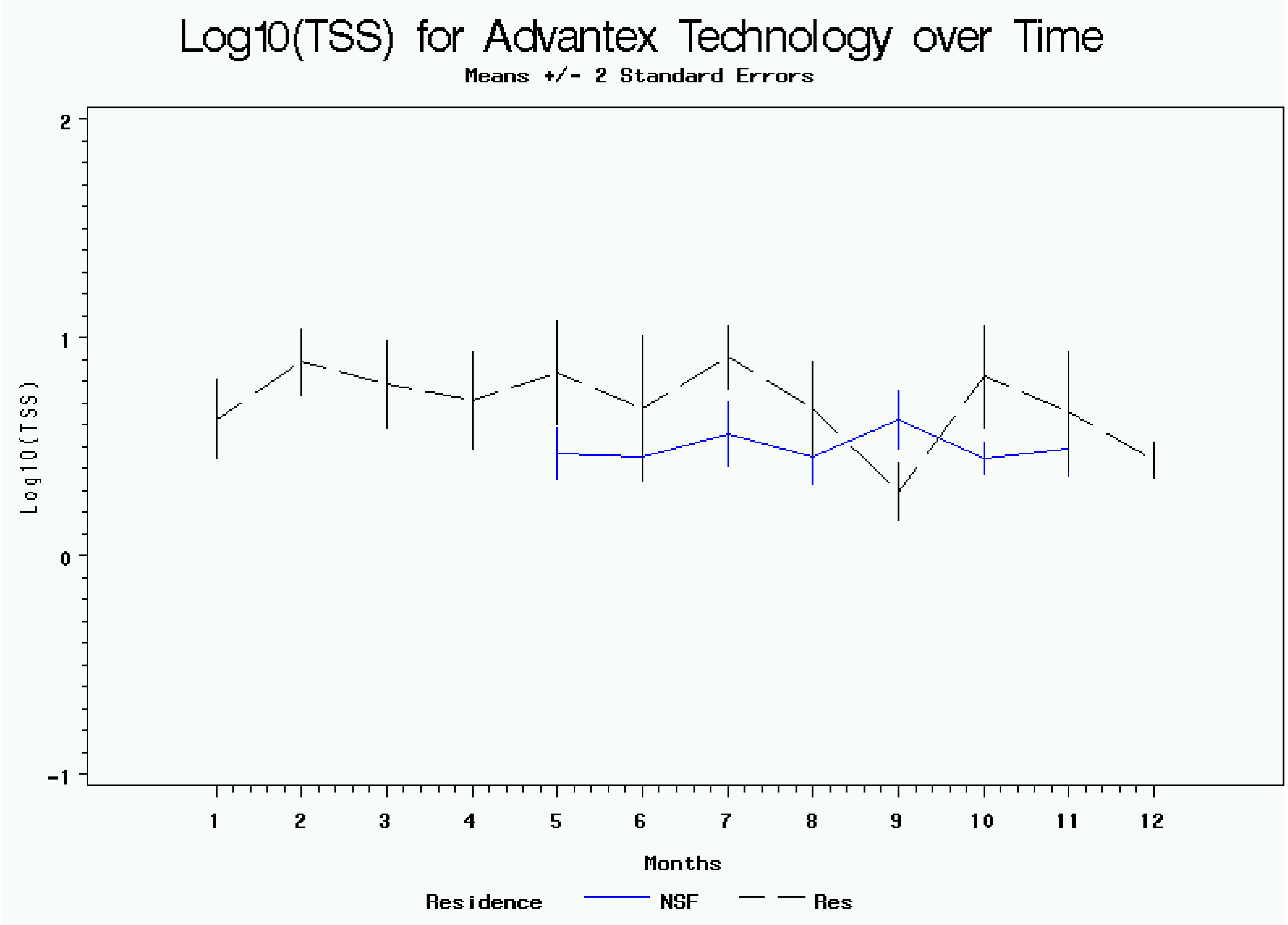


Figure 2.18.b. Monthly Means for BOD: Advantex Technology.

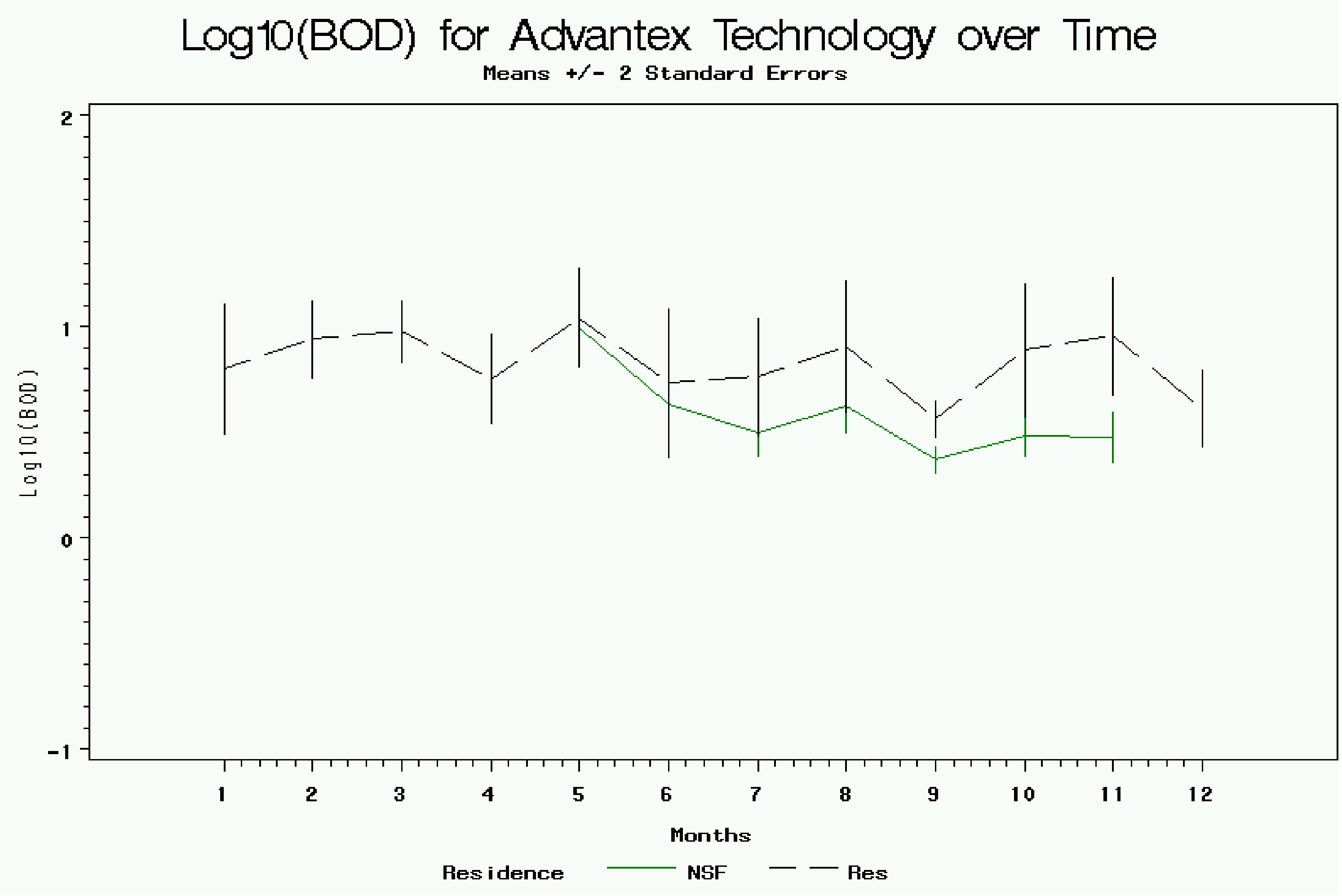


Figure 2.19.a. Monthly Means for TSS: Bioclere Technology.

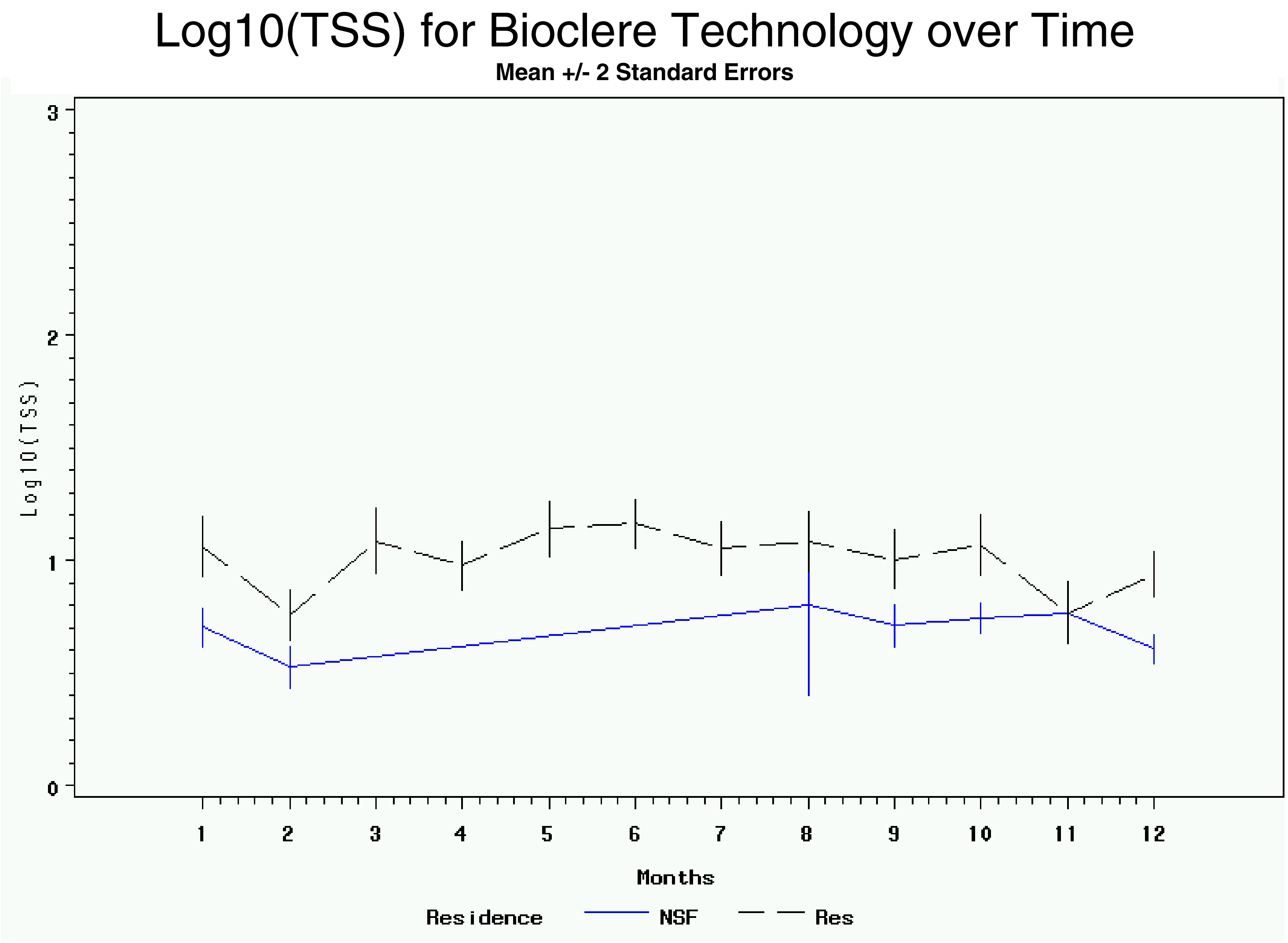


Figure 2.19.b. Monthly Means for BOD: Bioclere Technology.

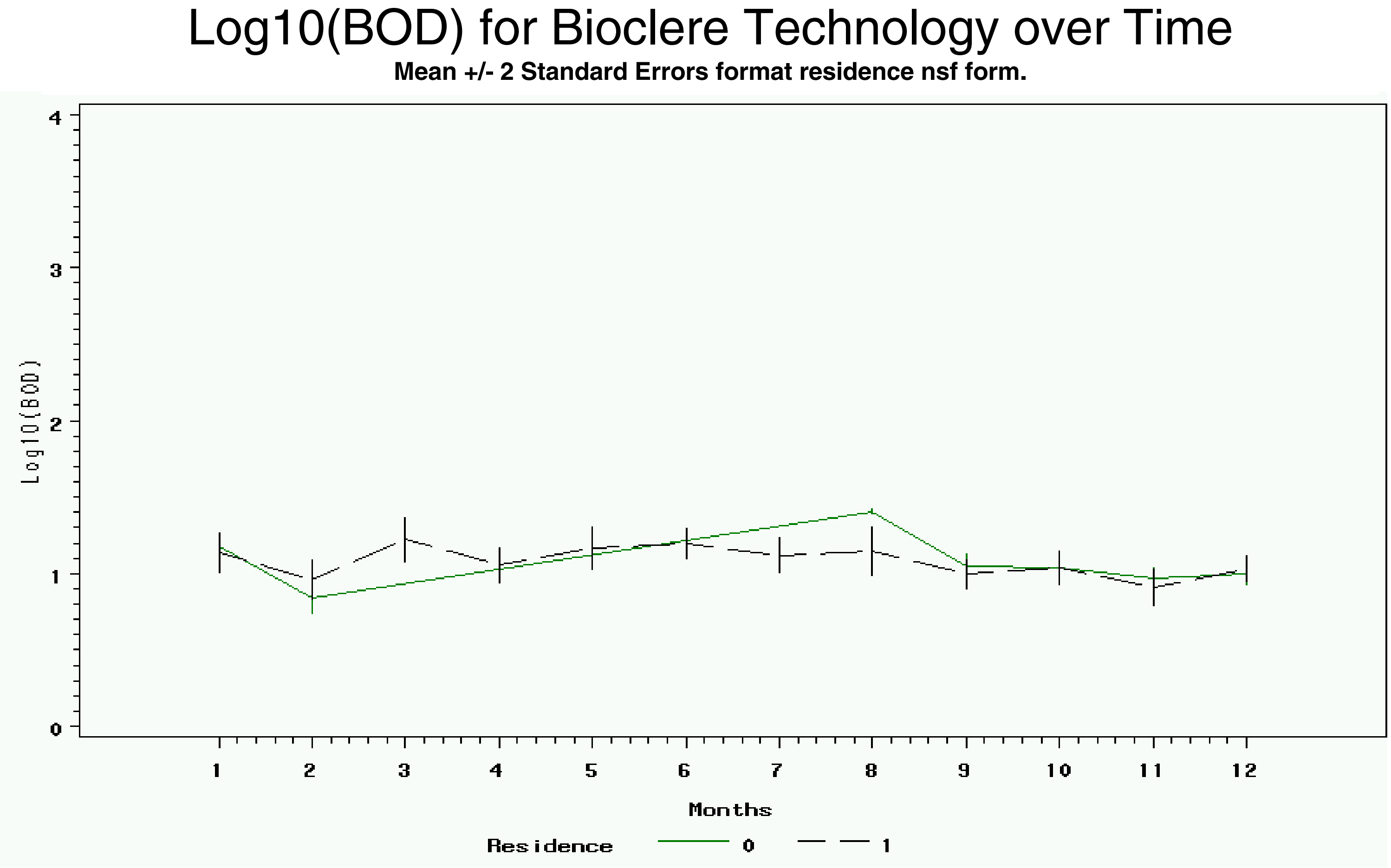


Figure 2.20.a. Monthly Means for TSS: FAST Technology.

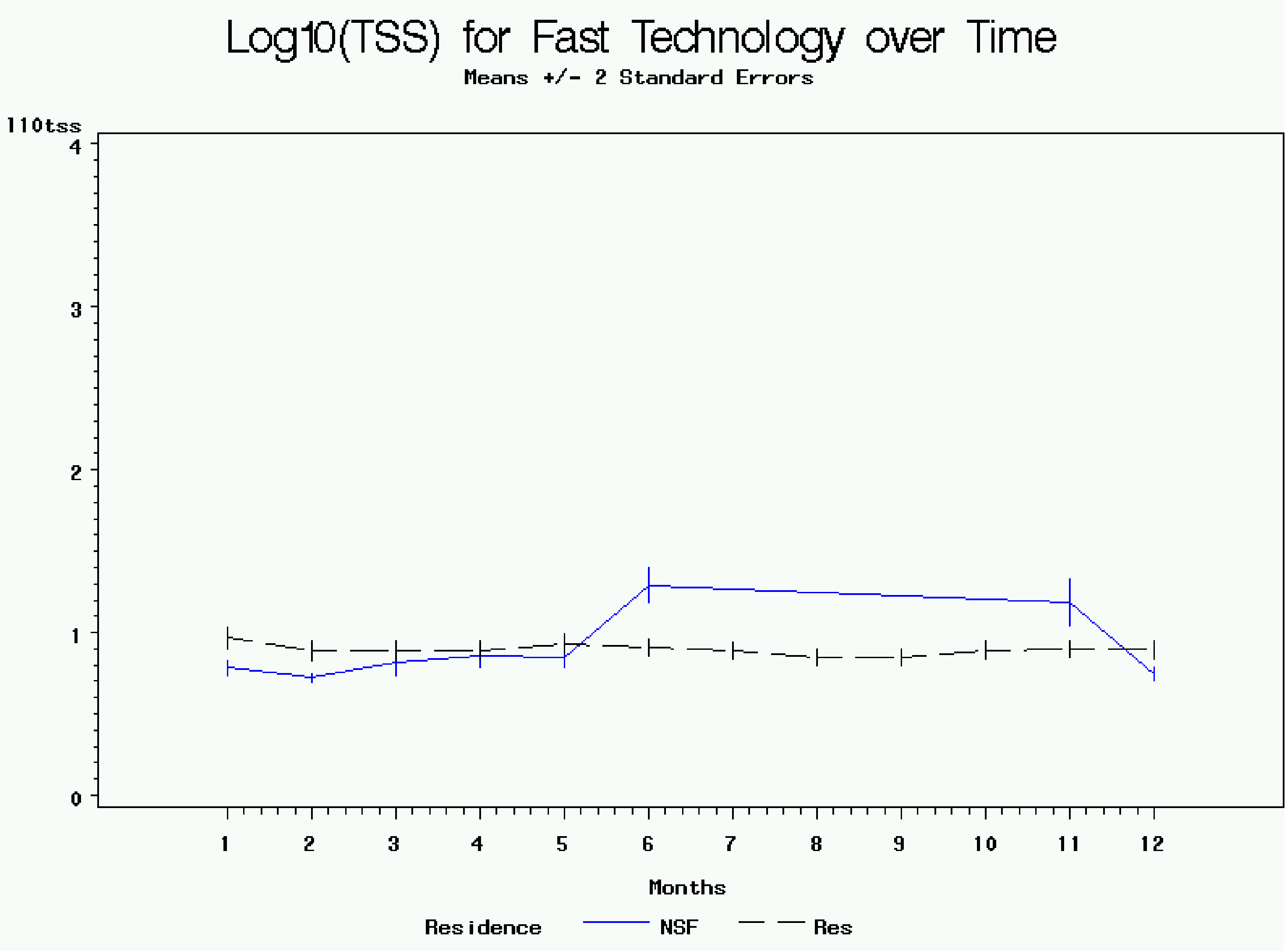


Figure 2.20.b. Monthly Means for BOD: FAST Technology.

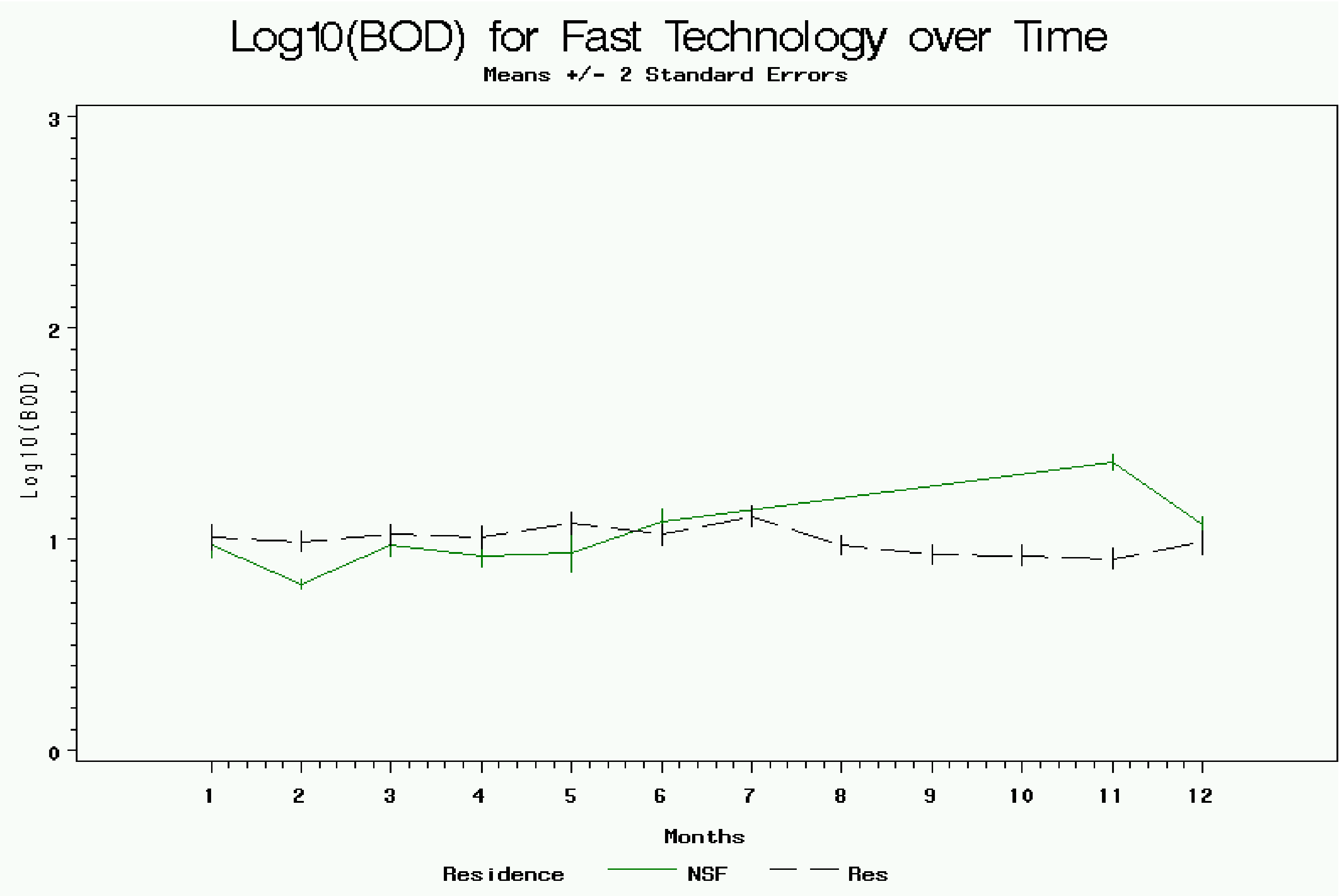


Figure 2.21.a. Monthly Medians for TSS: Advantex Technology.

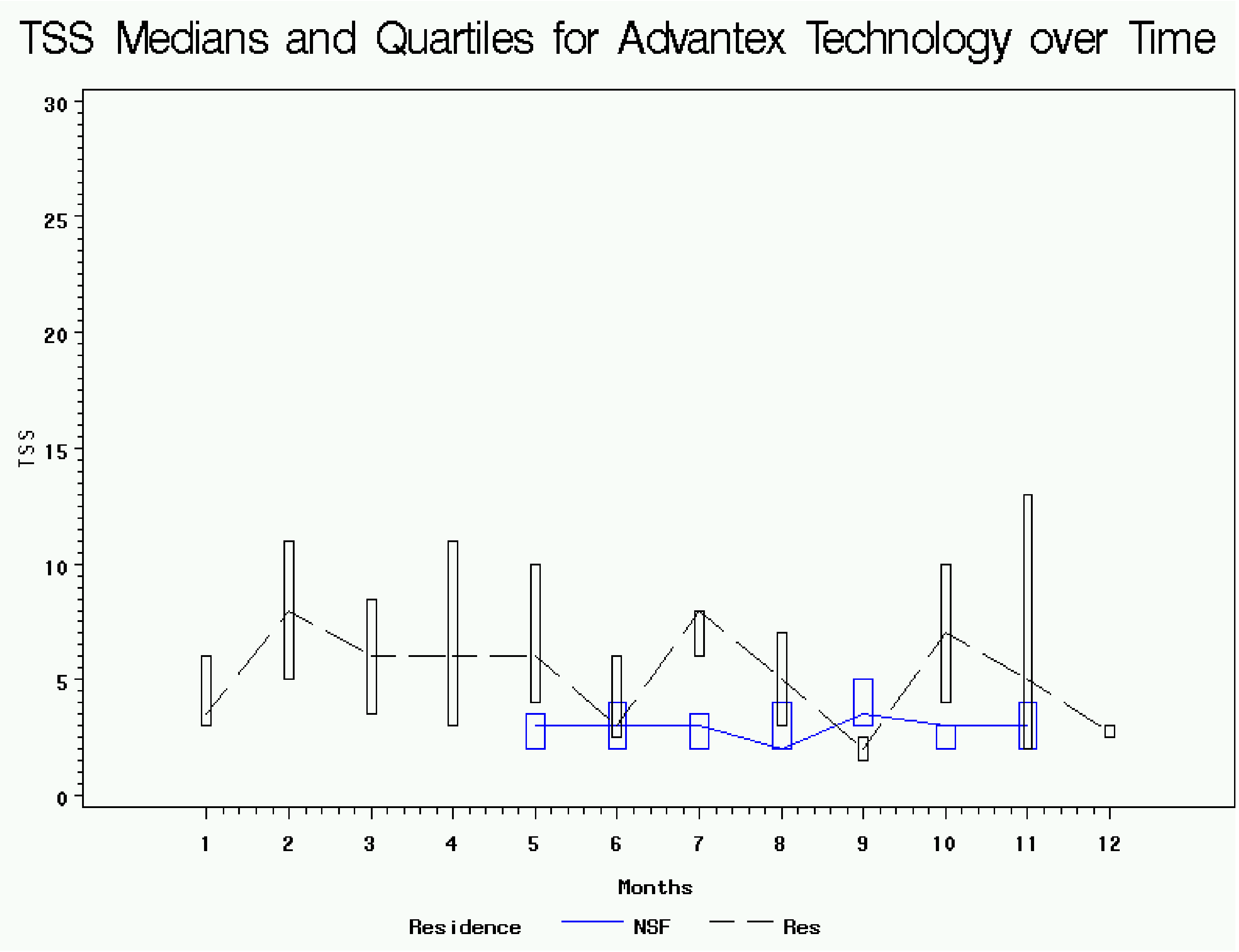


Figure 2.21.b. Monthly Medians for BOD: Advantex Technology.

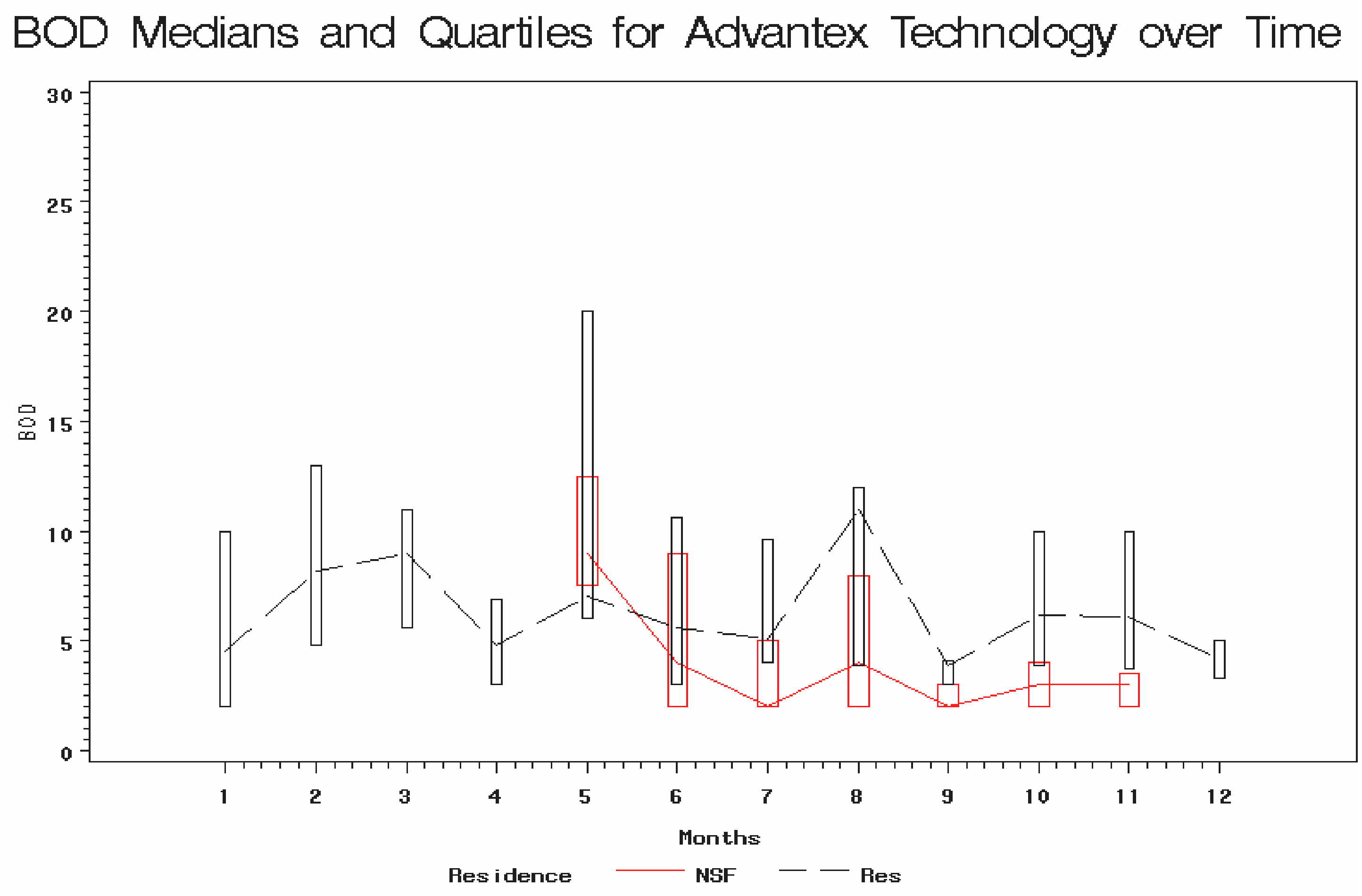


Figure 2.22.a. Monthly Medians for TSS: Bioclere Technology.

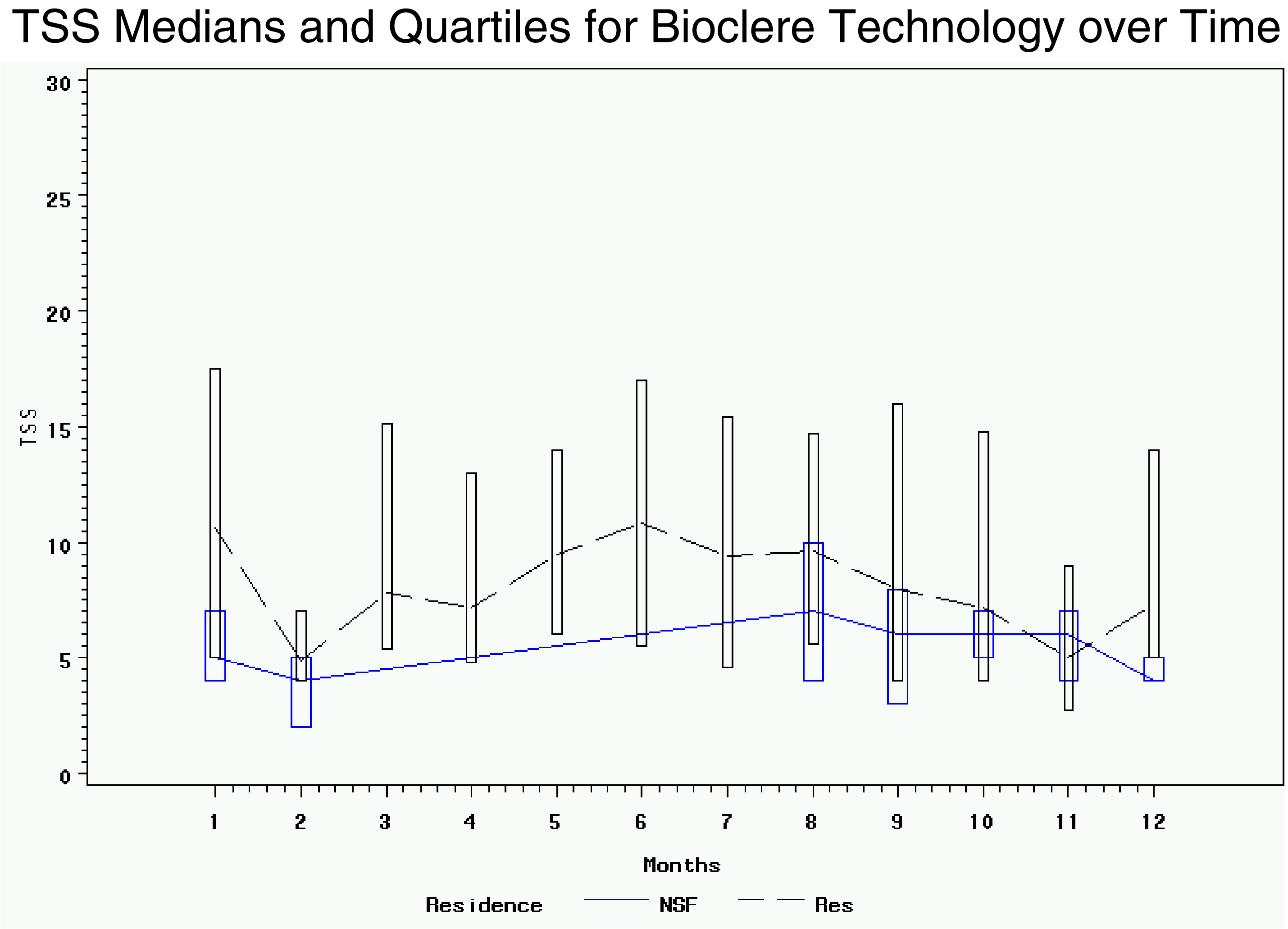


Figure 2.22.b. Monthly Medians for BOD: Bioclere Technology.

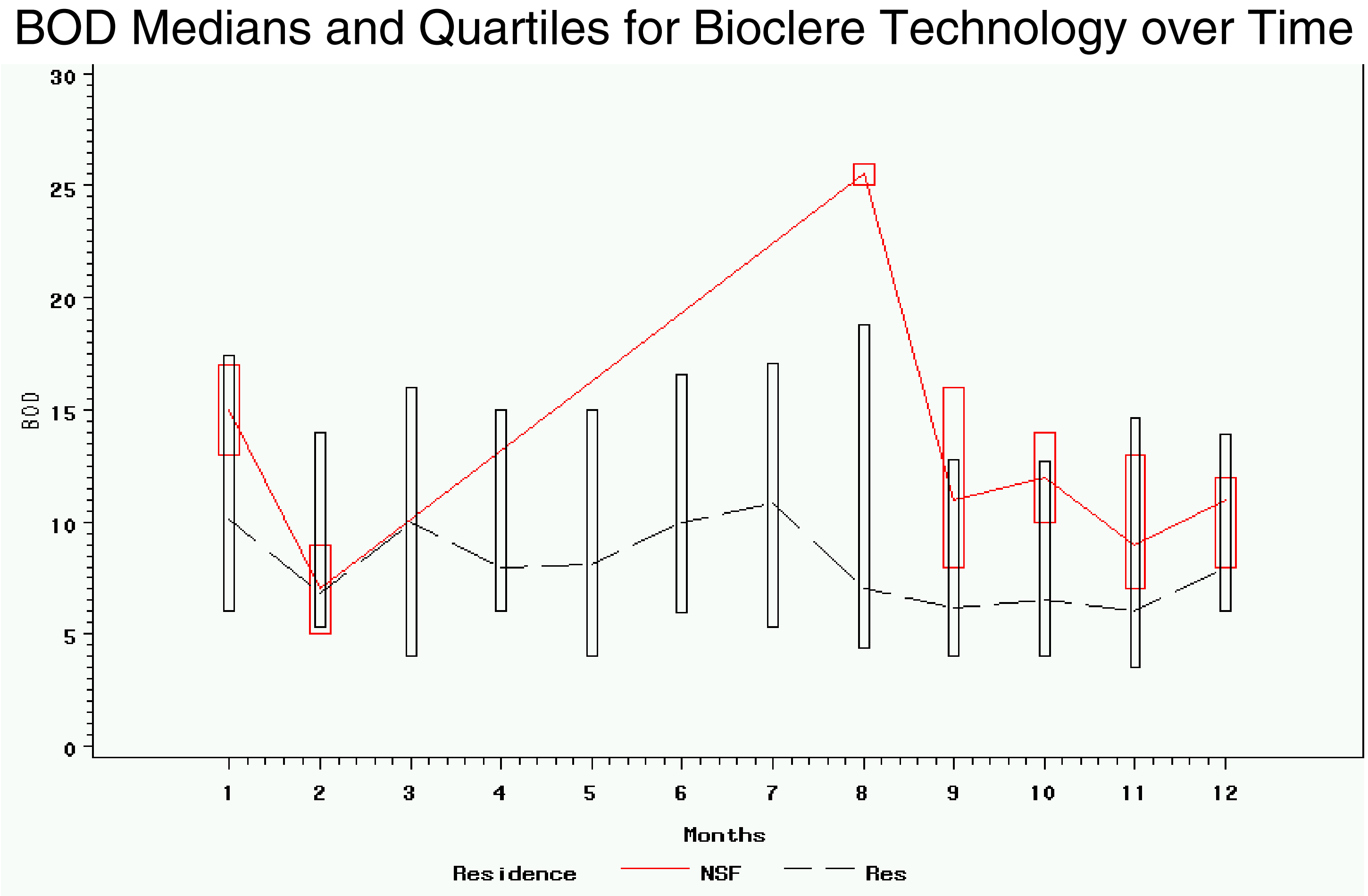


Figure 2.23.a. Monthly Medians for TSS: FAST Technology.

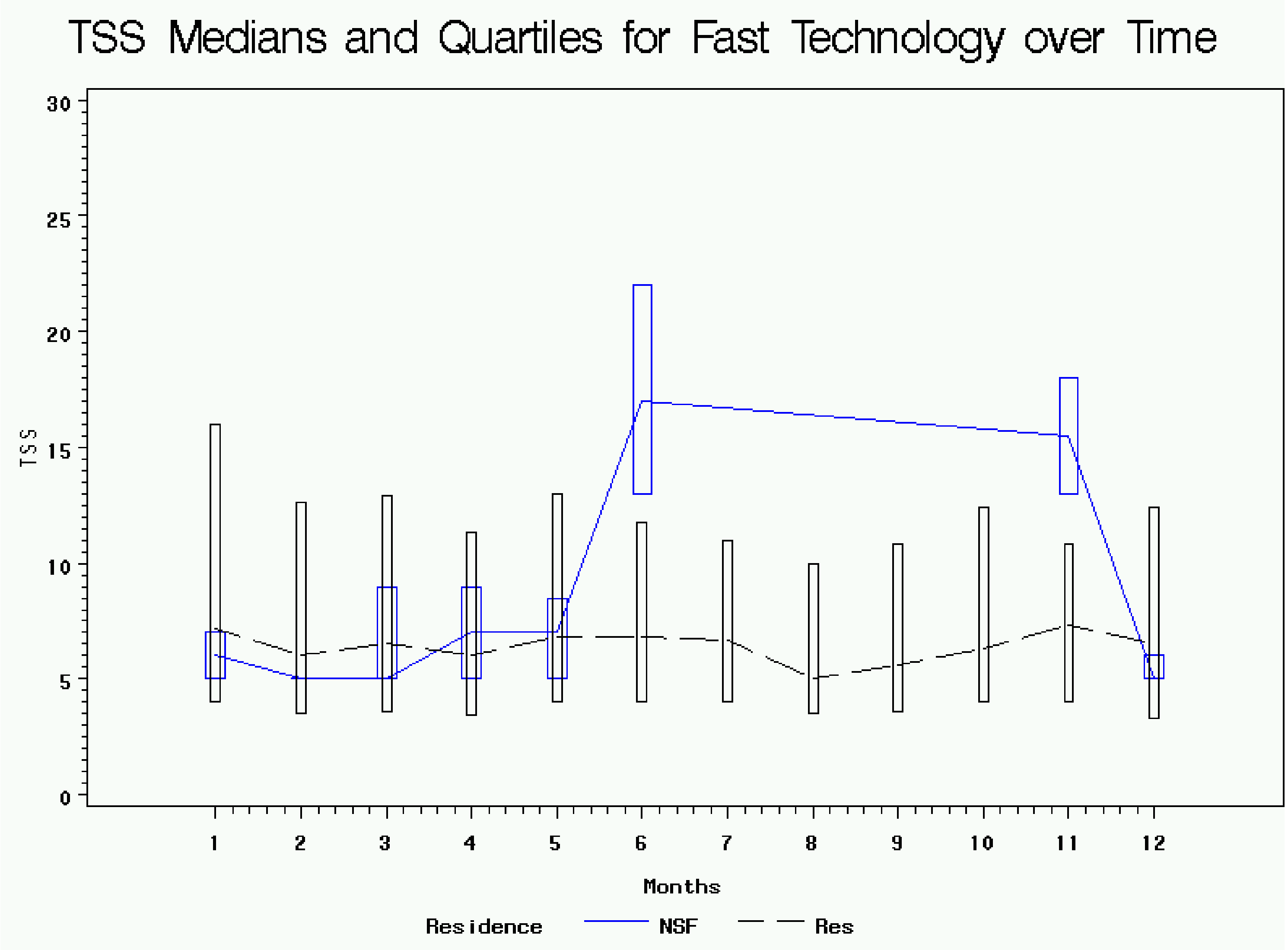


Figure 2.23.b. Monthly Medians for BOD: FAST Technology.

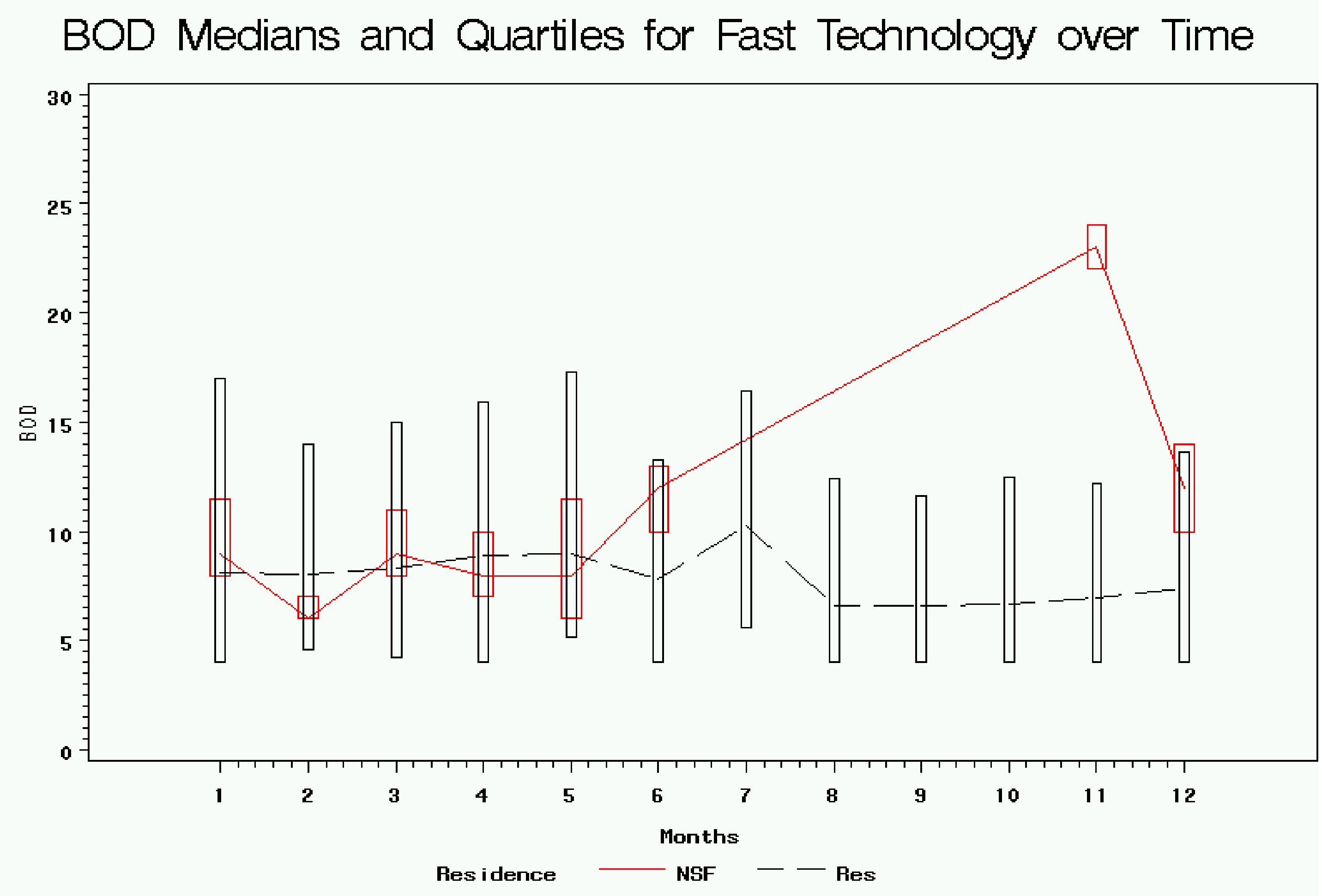


Figure 2.24. Standard Error of the Mean: Advantex TSS: Variance Between= 49.1: Variance Within= 7.2

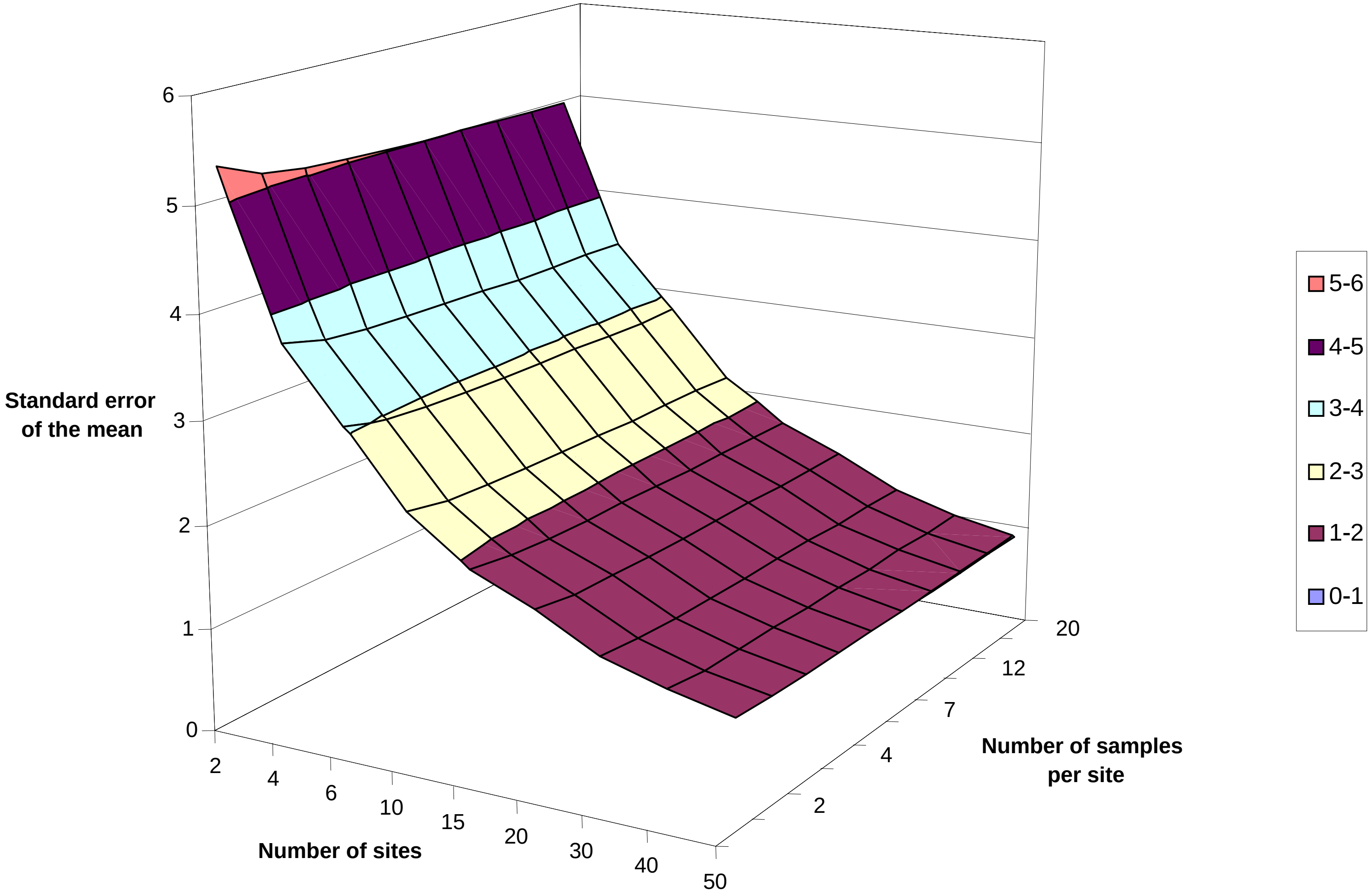


Figure 2.25. Standard Error of the Mean: Advantex TSS Variance Between= 49.1: Variance Within= 7.2

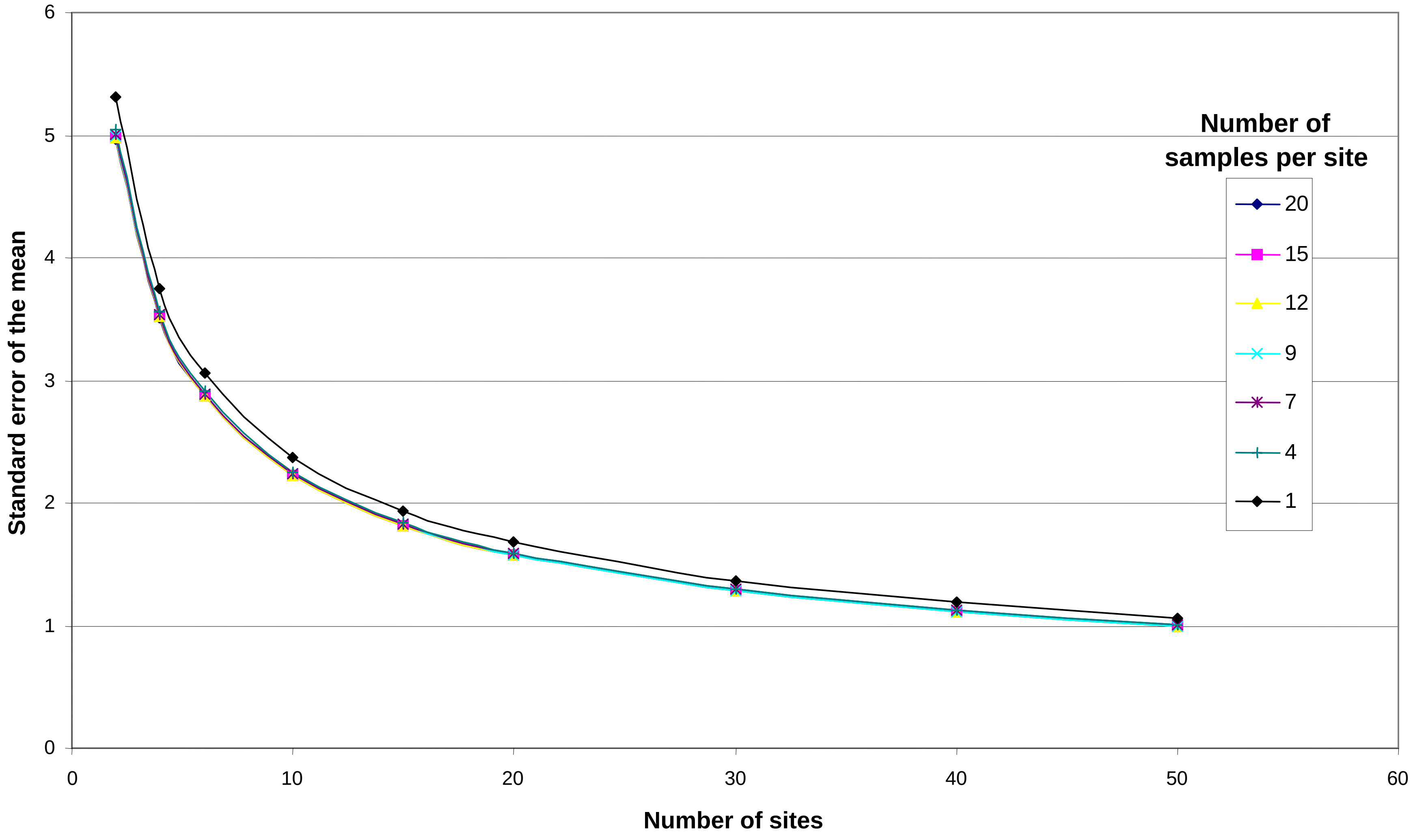


Fig 2.26. Sample Sizes for 84, 95 and 99% Constant Width Confidence Intervals Standard Error of the Mean: Advantex TSS: Variance Between= 49.1: Variance Within= 7.2

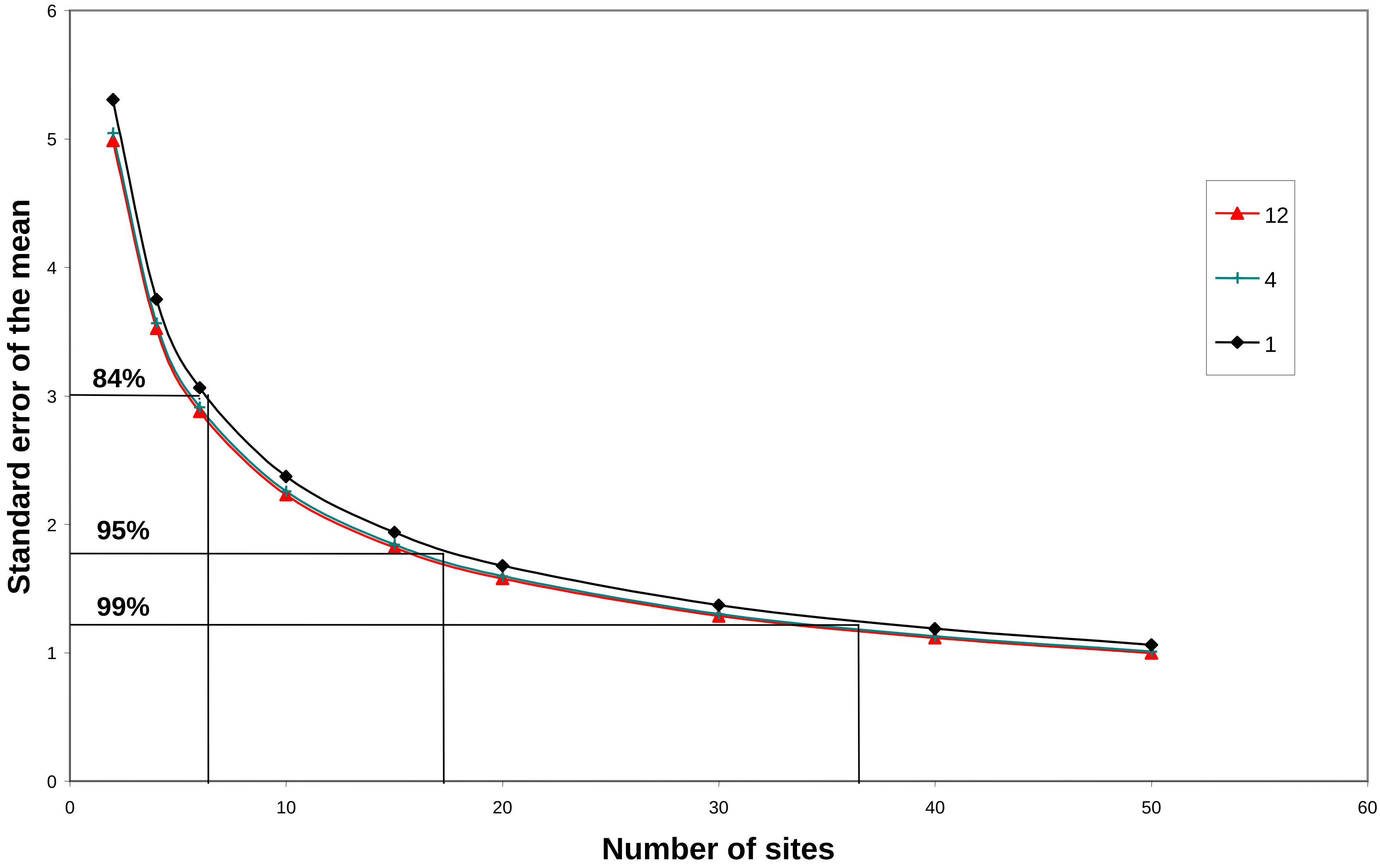


Figure 2.27. Bioclere: BOD NSF and 'New' Test Data

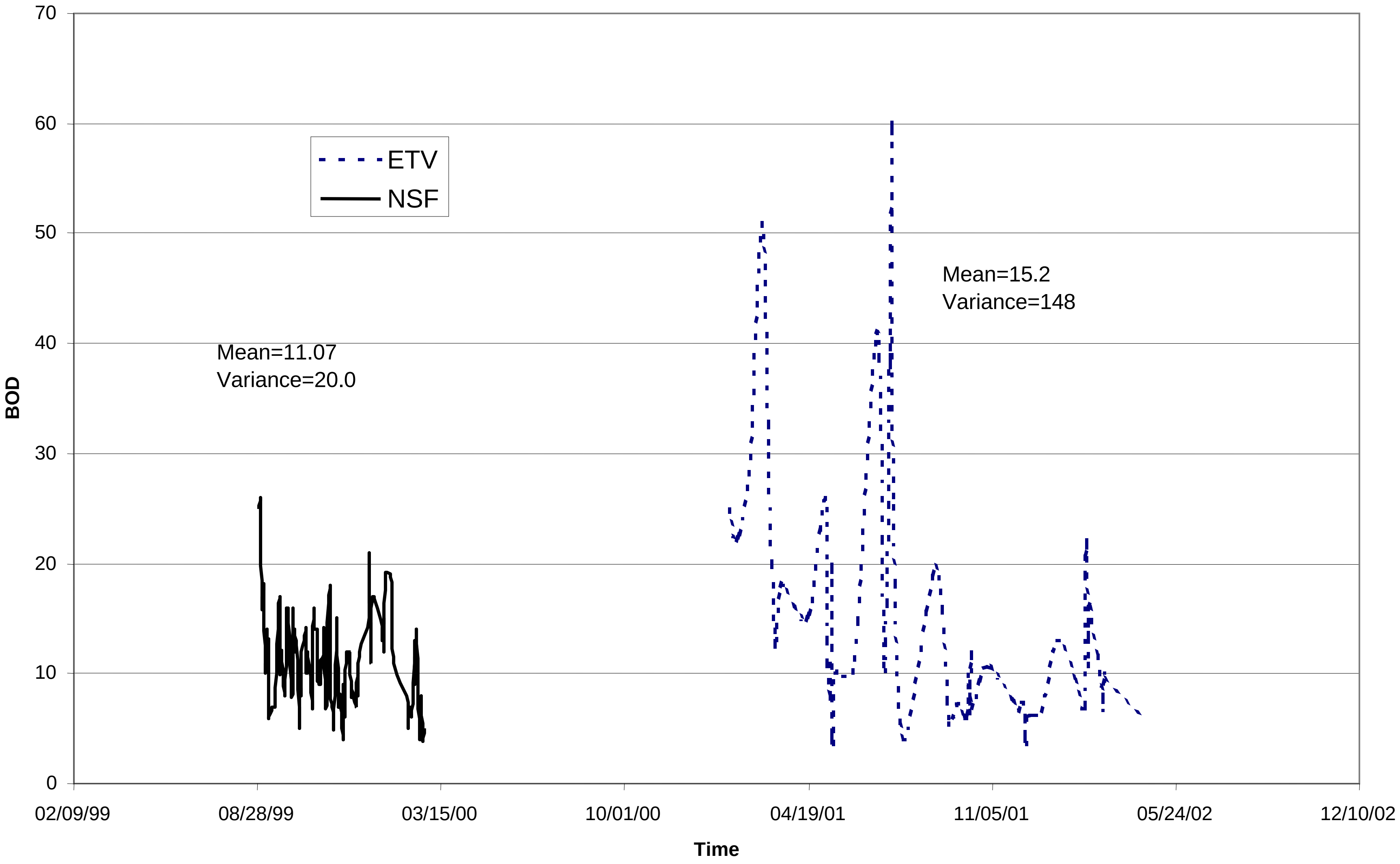
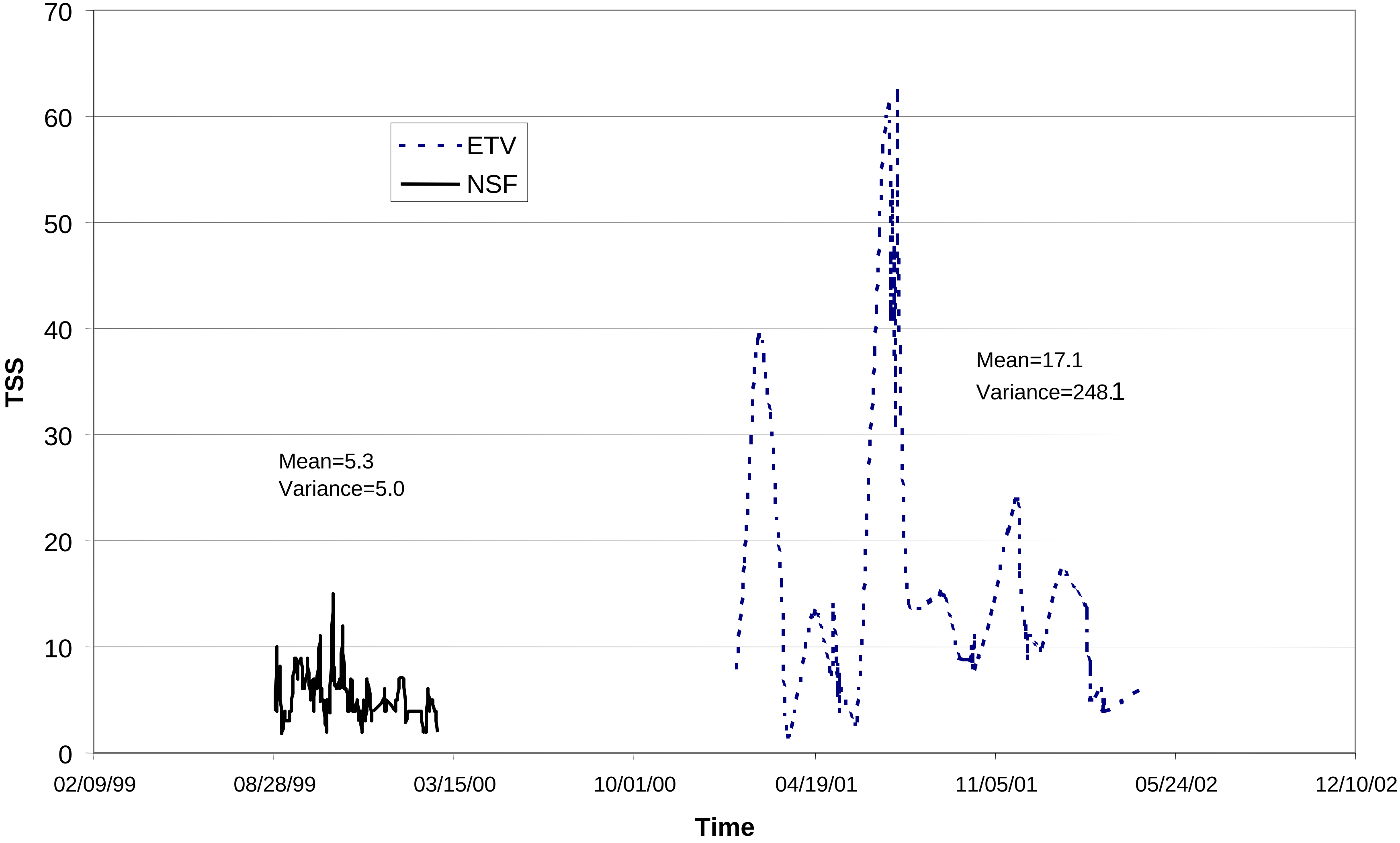


Figure 2.28. Bioclere: TSS NSF and 'New' Test Data



B FIGURES FOR OTHER COMBINATIONS OF TECHNOLOGY AND VARIABLES FOR BOTH UNTRANSFORMED AND LOG10 TRANSFORMED DATA

Figure B-1. Untransformed Data. Standard Error of the Mean: Advantex BOD: Variance Between = 100.9: Variance Within = 29.1

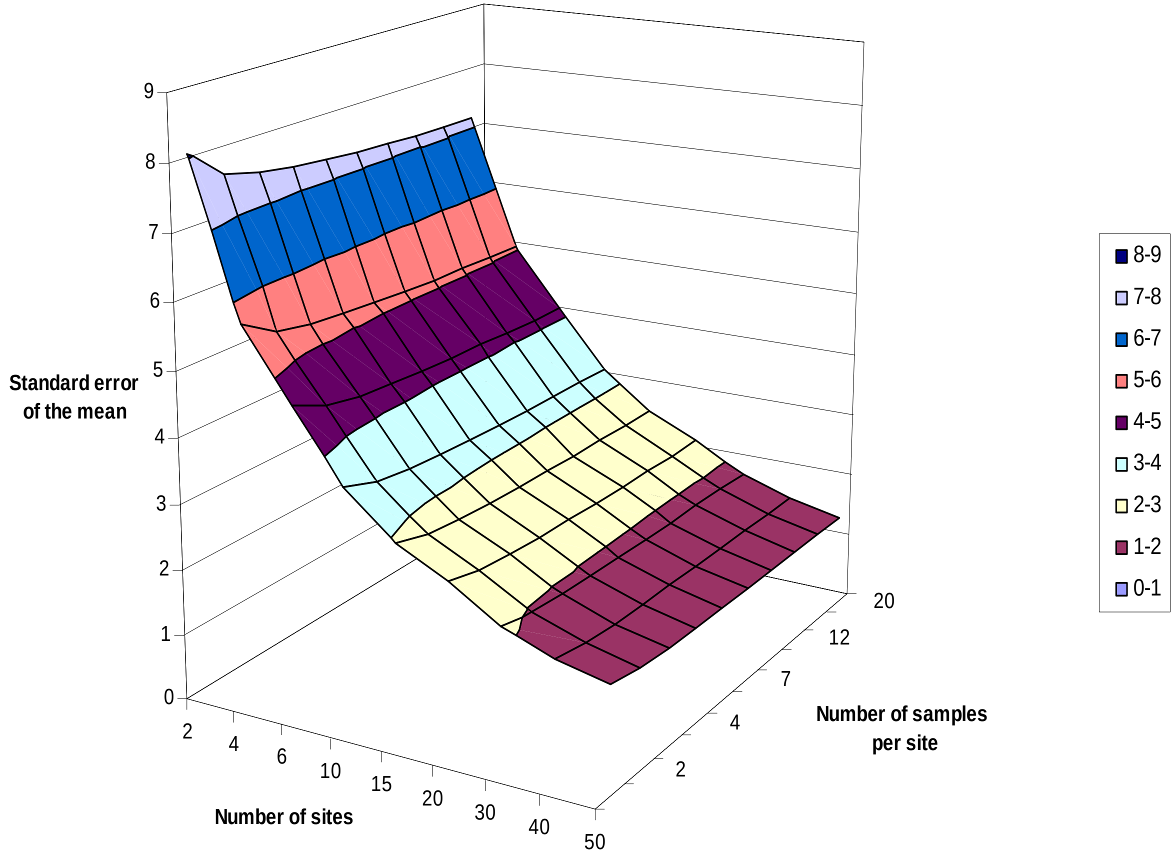


Figure B-2. Untransformed Data. Standard Error of the Mean: Advantex BOD: Variance Between = 100.9: Variance Within = 29.1

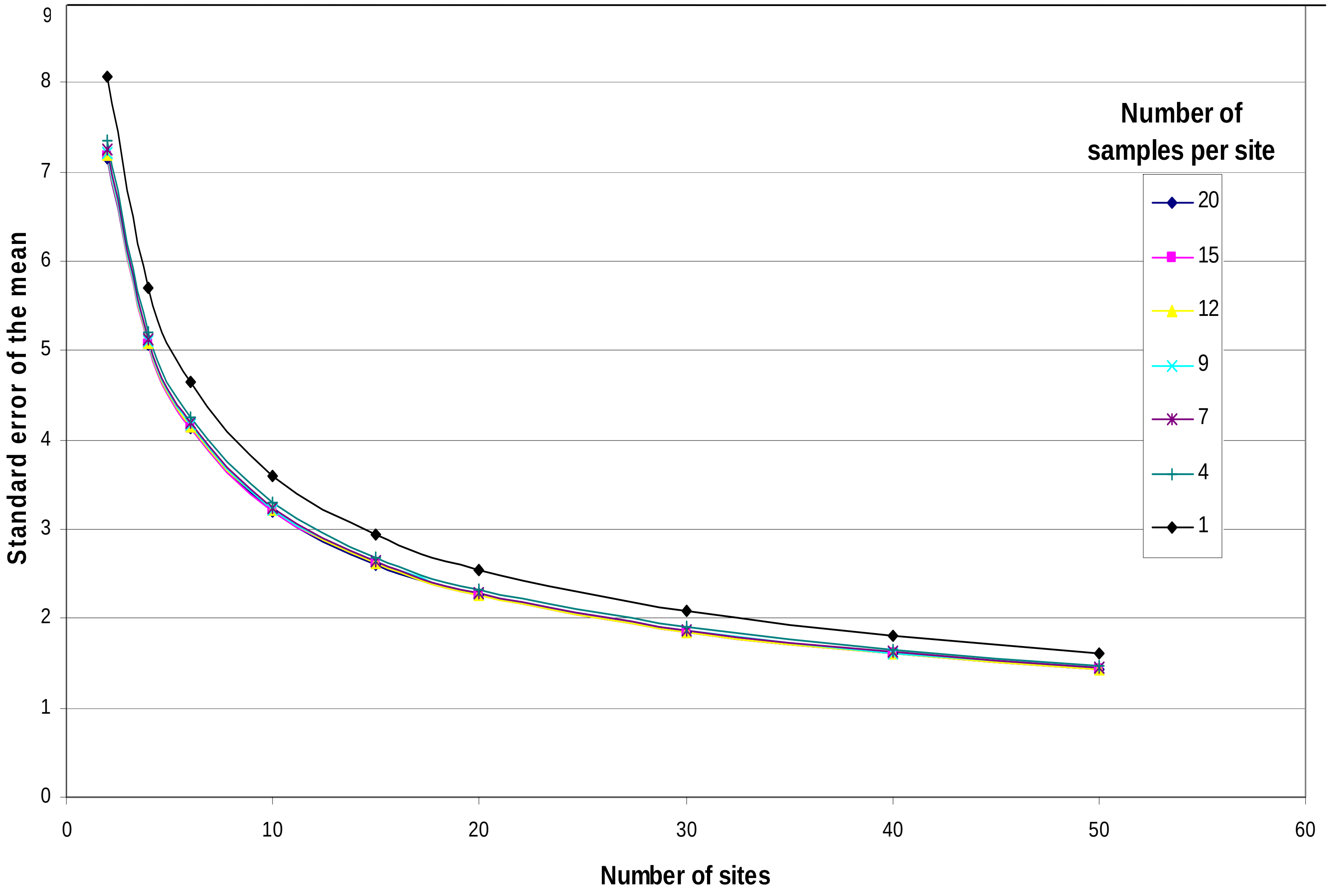


Figure B-3. Untransformed Data. Standard Error of the Mean: Bioclere TSS: Variance Between = 1946.1: Variance Within = 154.9

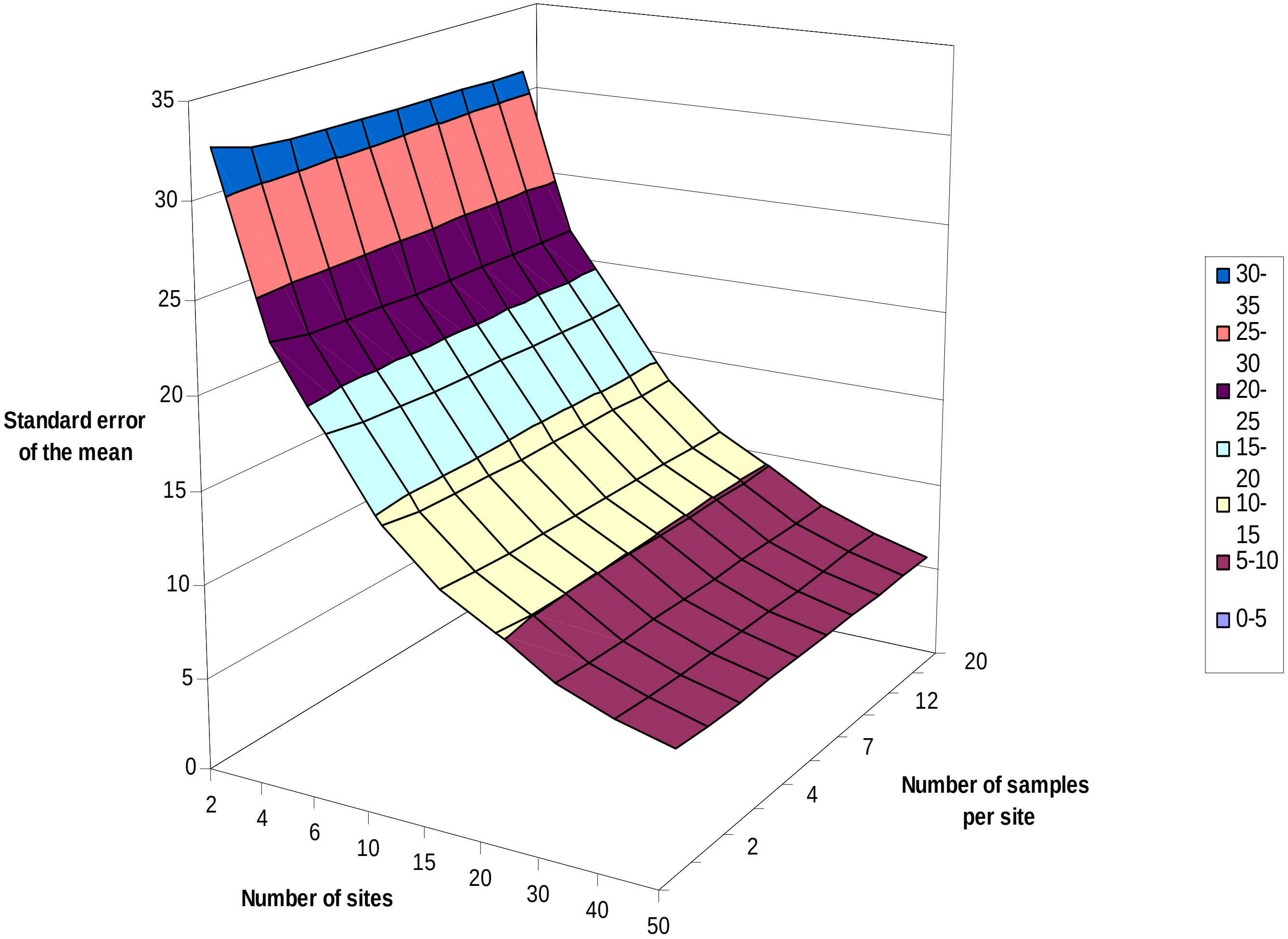


Figure B-4. Untransformed Data. Standard Error of the Mean: Bioclere TSS: Variance Between = 1946.1: Variance Within = 154.9

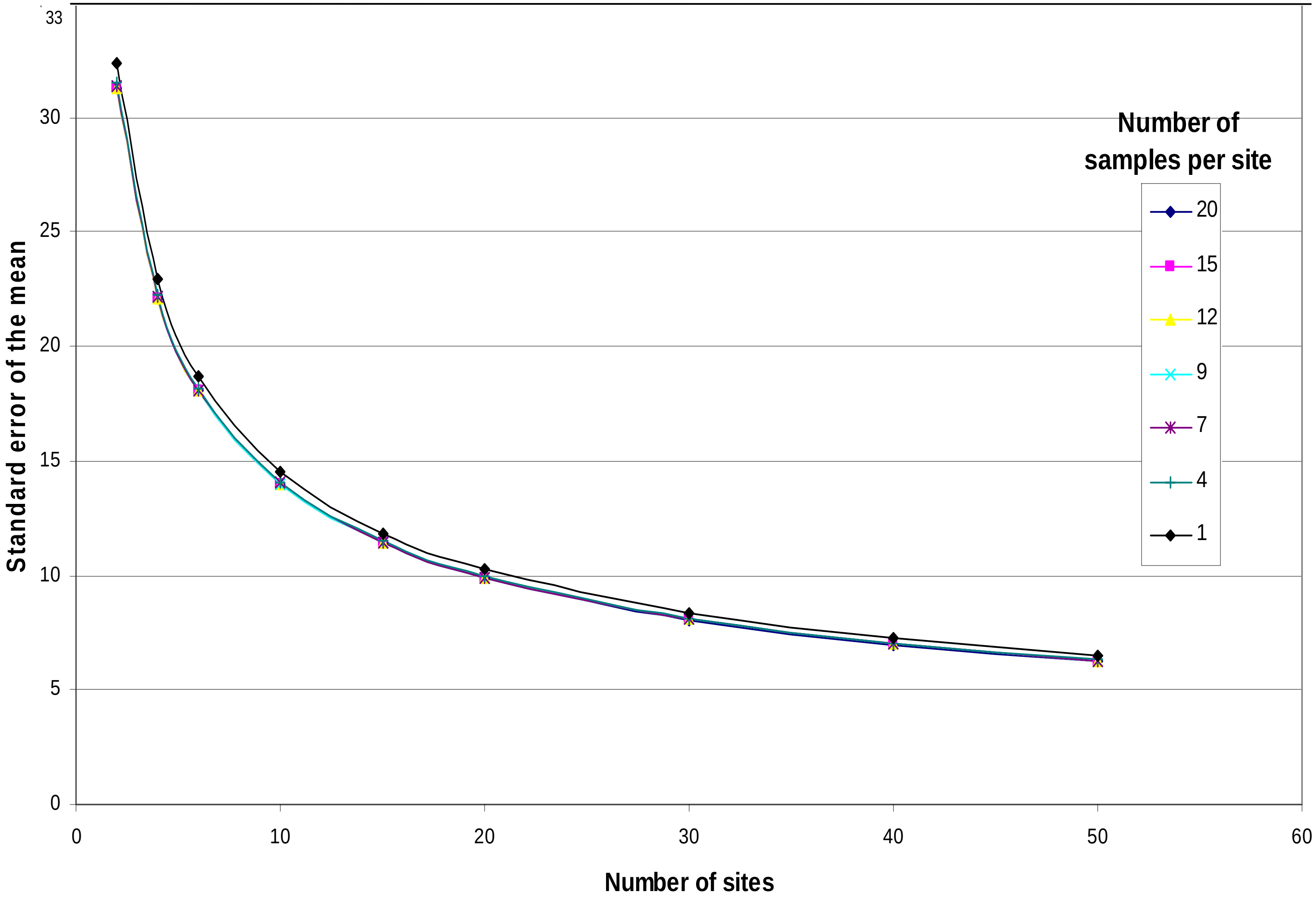


Figure B-5. Untransformed Data. Standard Error of the Mean: Bioclere BOD: Variance Between = 2286.6: Variance Within = 272.2

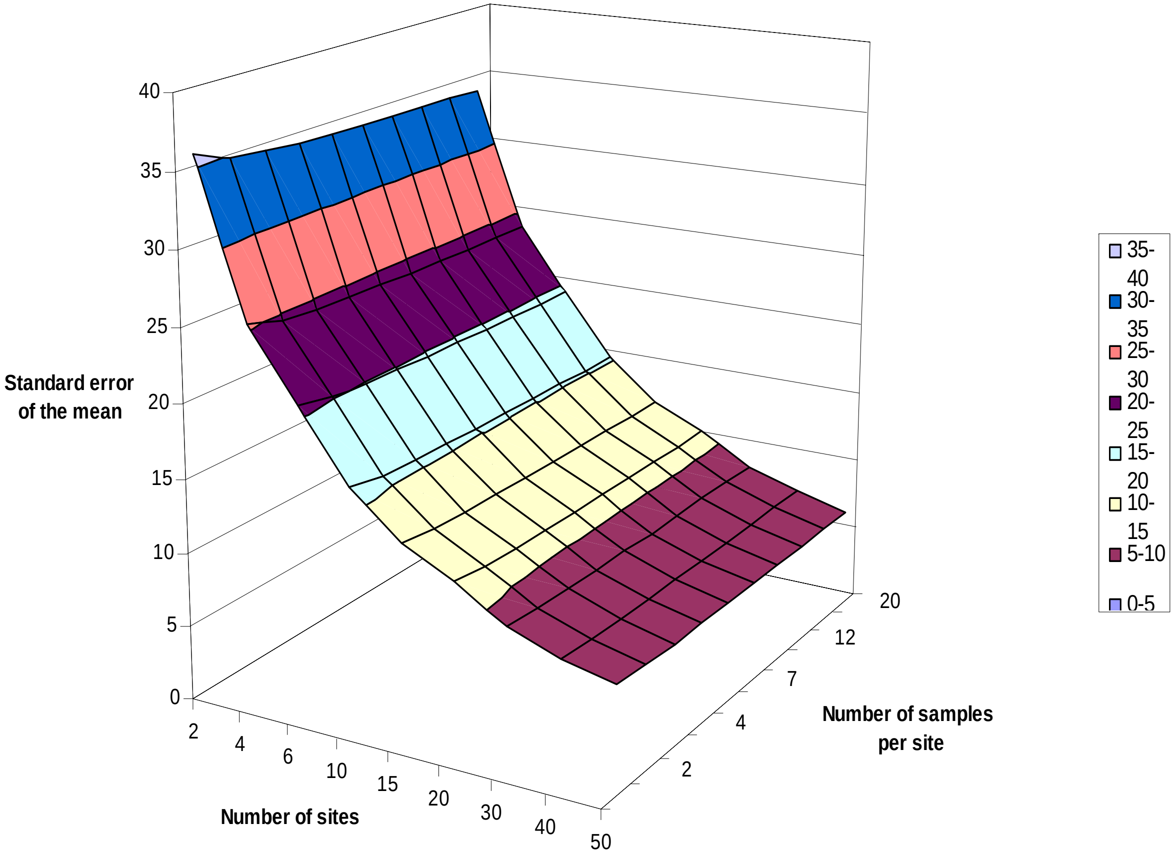


Figure B-6. Untransformed Data. Standard Error of the Mean: Bioclere BOD: Variance Between = 2286.6: Variance Within = 272.2

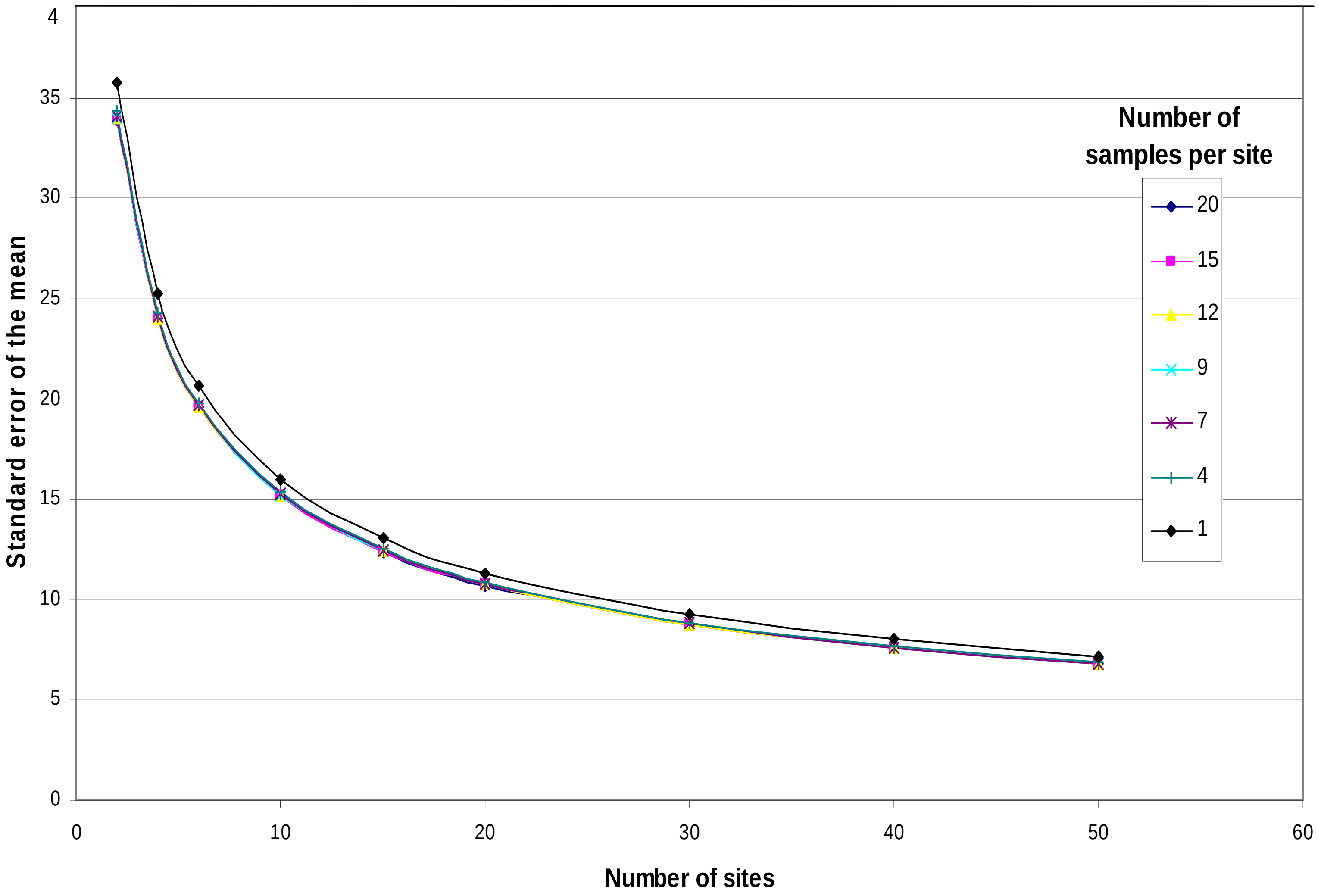


Figure B-7. Untransformed Data. Standard Error of the Mean: FAST TSS: Variance Between = 2484.3: Variance Within = 82.5

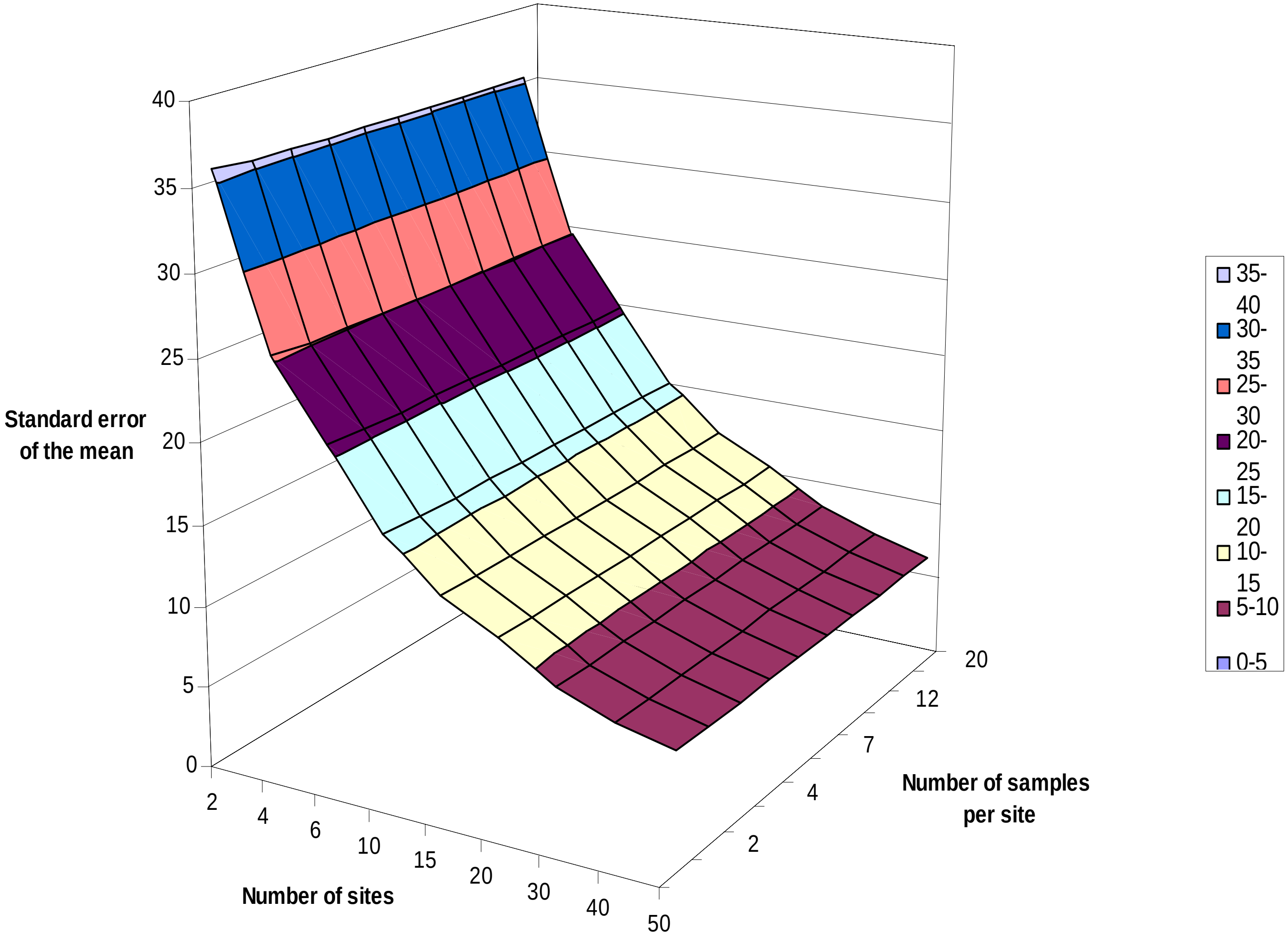


Figure B-8. Untransformed Data. Standard Error of the Mean: FAST TSS: Variance Between = 2484.3: Variance Within = 82.5

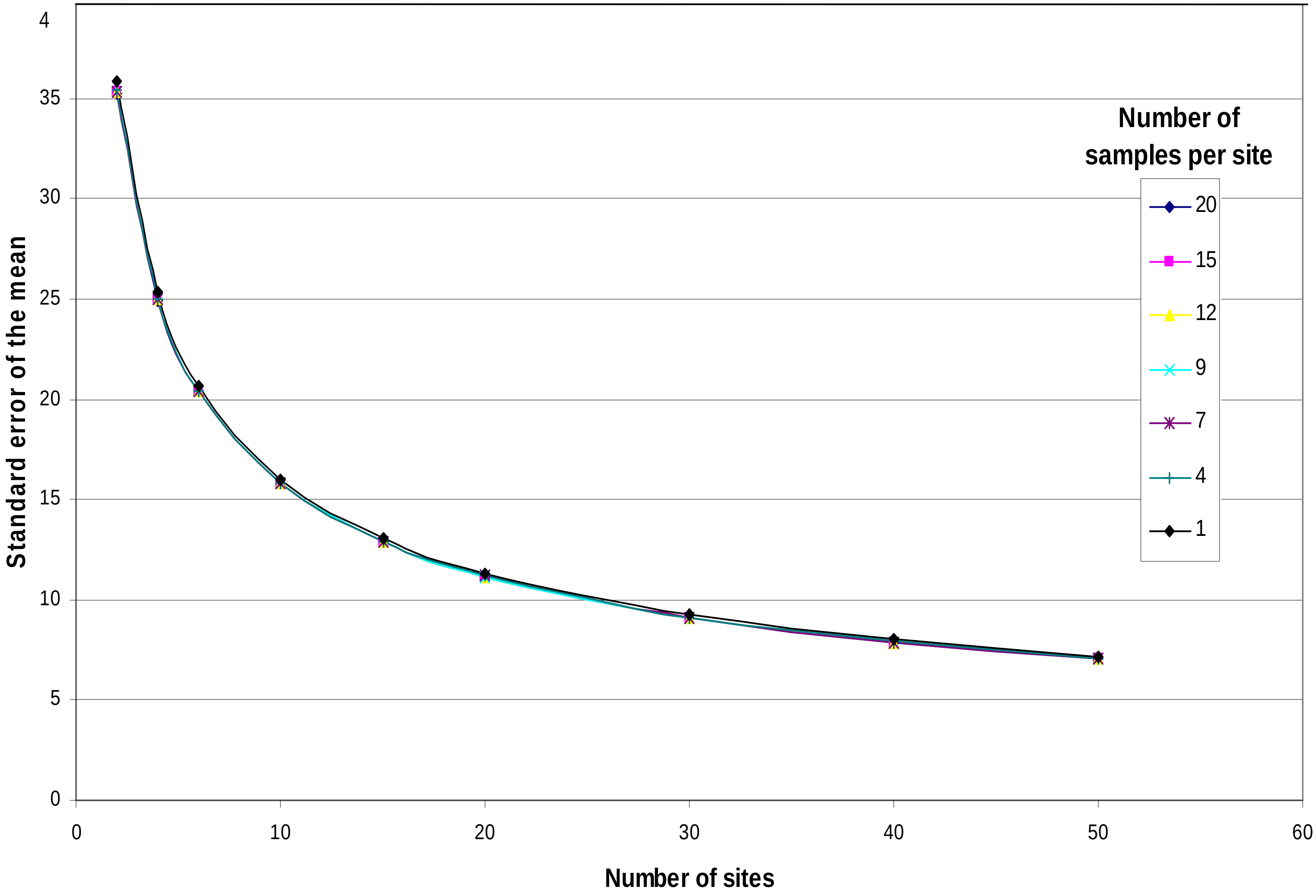


Figure B-9. Untransformed Data. Standard Error of the Mean: FAST BOD: Variance Between = 374.8: Variance Within = 65

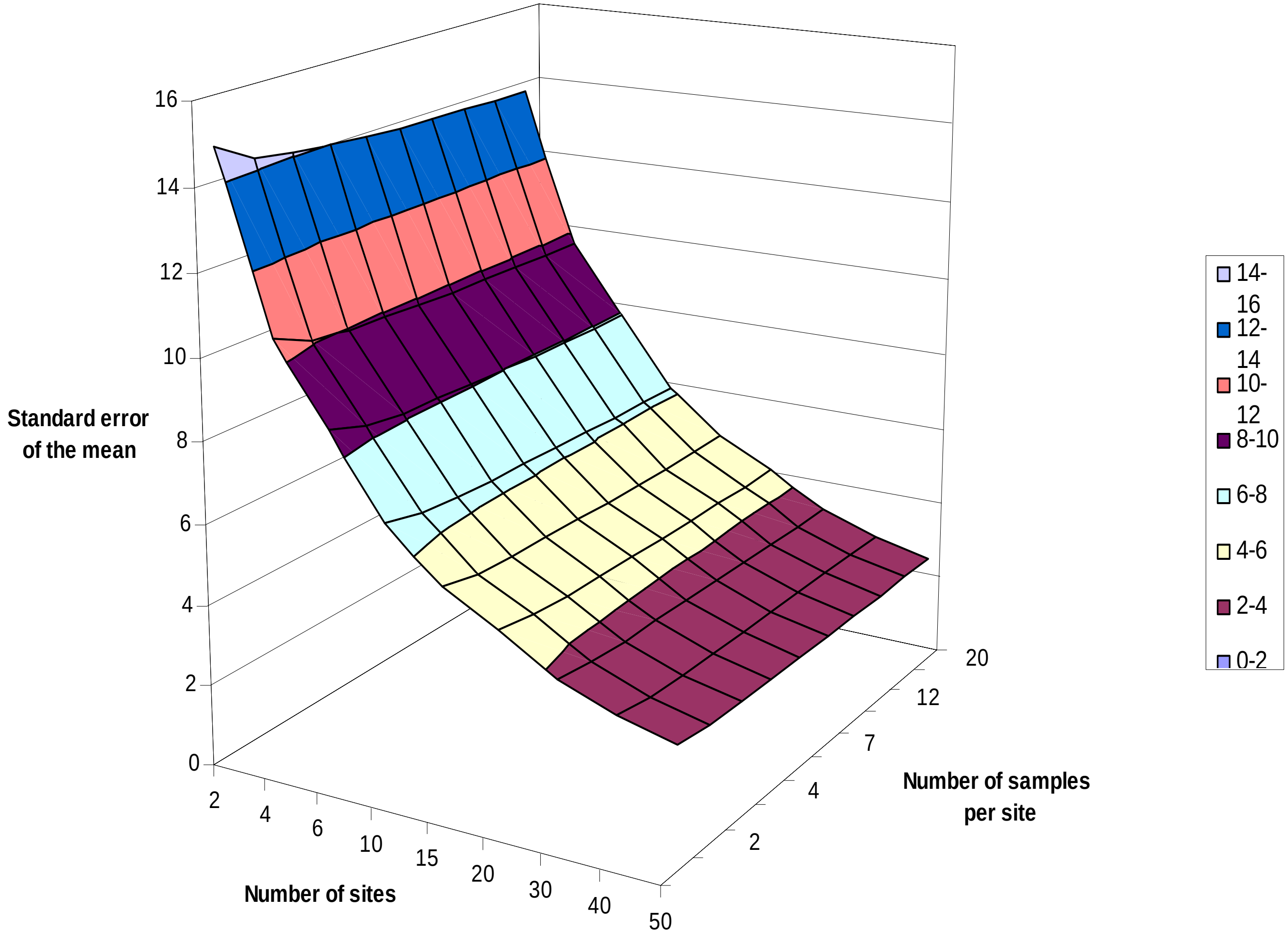


Figure B-10. Untransformed Data. Standard Error of the Mean: FAST BOD: Variance Between = 374.8: Variance Within = 65

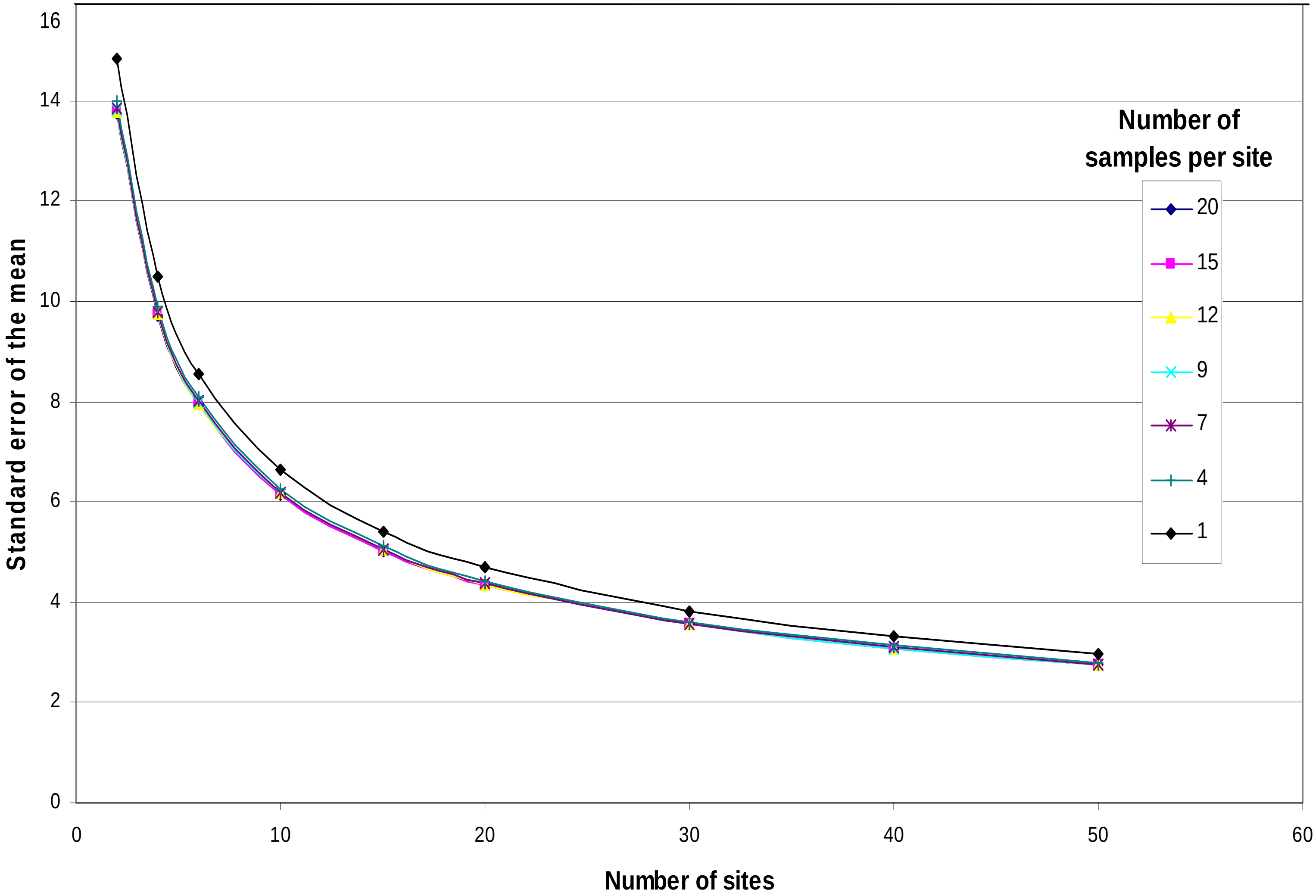


Figure B-11. Log10 Data. Standard Error of the Mean: Advantex TSS: Variance Between = 0.108: Variance Within = 0.017

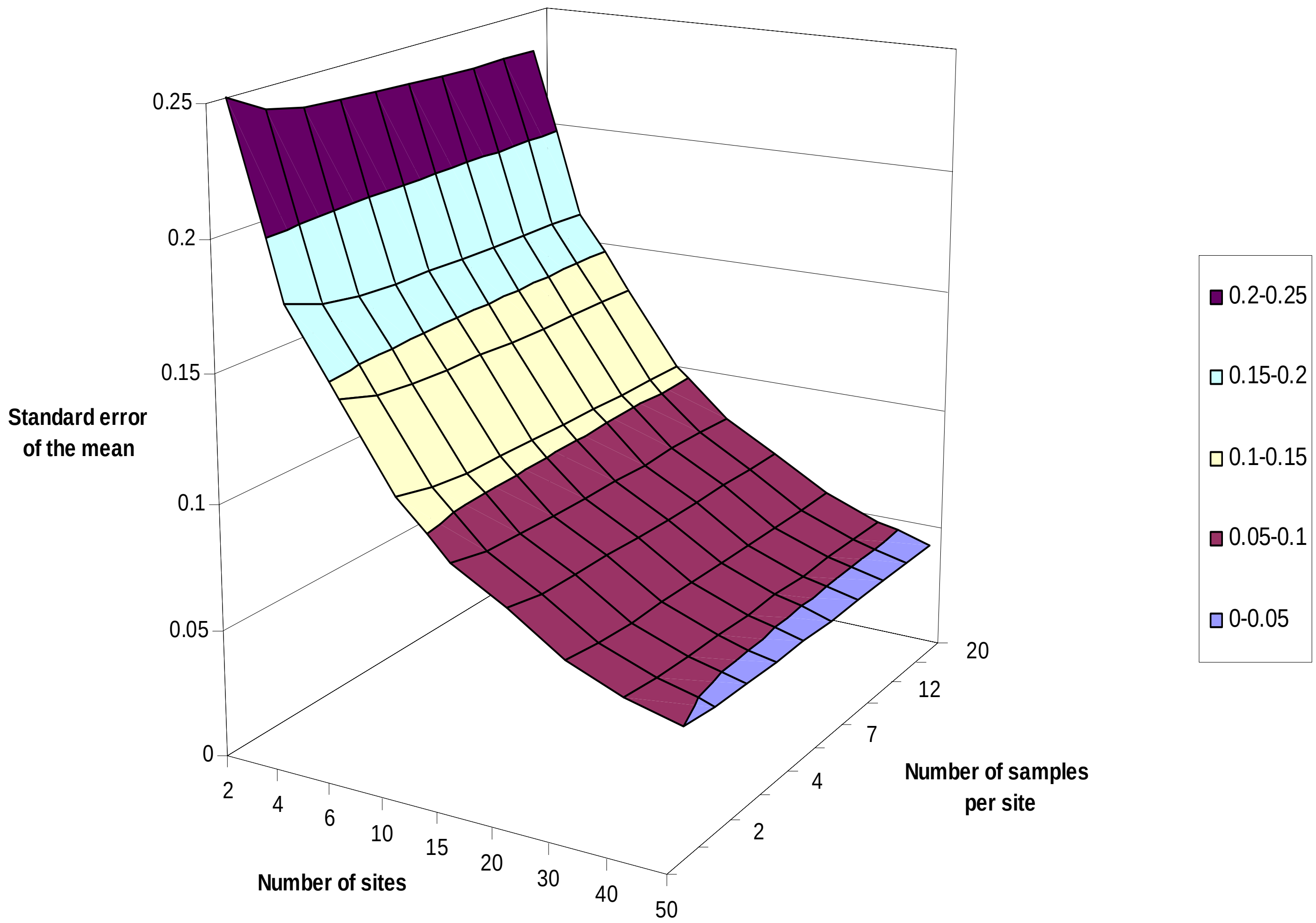


Figure B-12. Log10 Data. Standard Error of the Mean: Advantex TSS: Variance Between = 0.108: Variance Within = 0.017

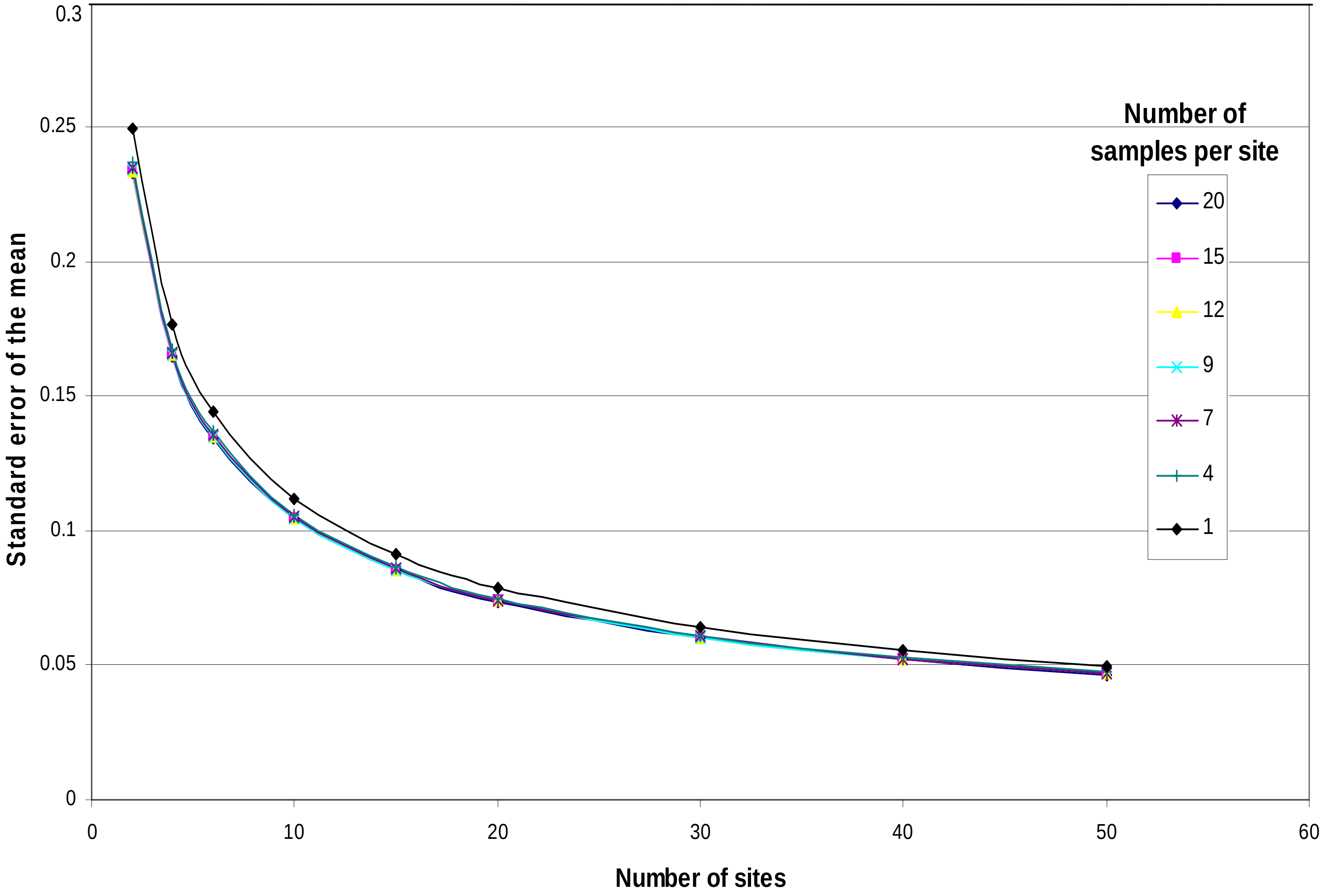


Figure B-13. Log10 Data. Standard Error of the Mean: Advantex BOD: Variance Between = 0.054: Variance Within = 0.047

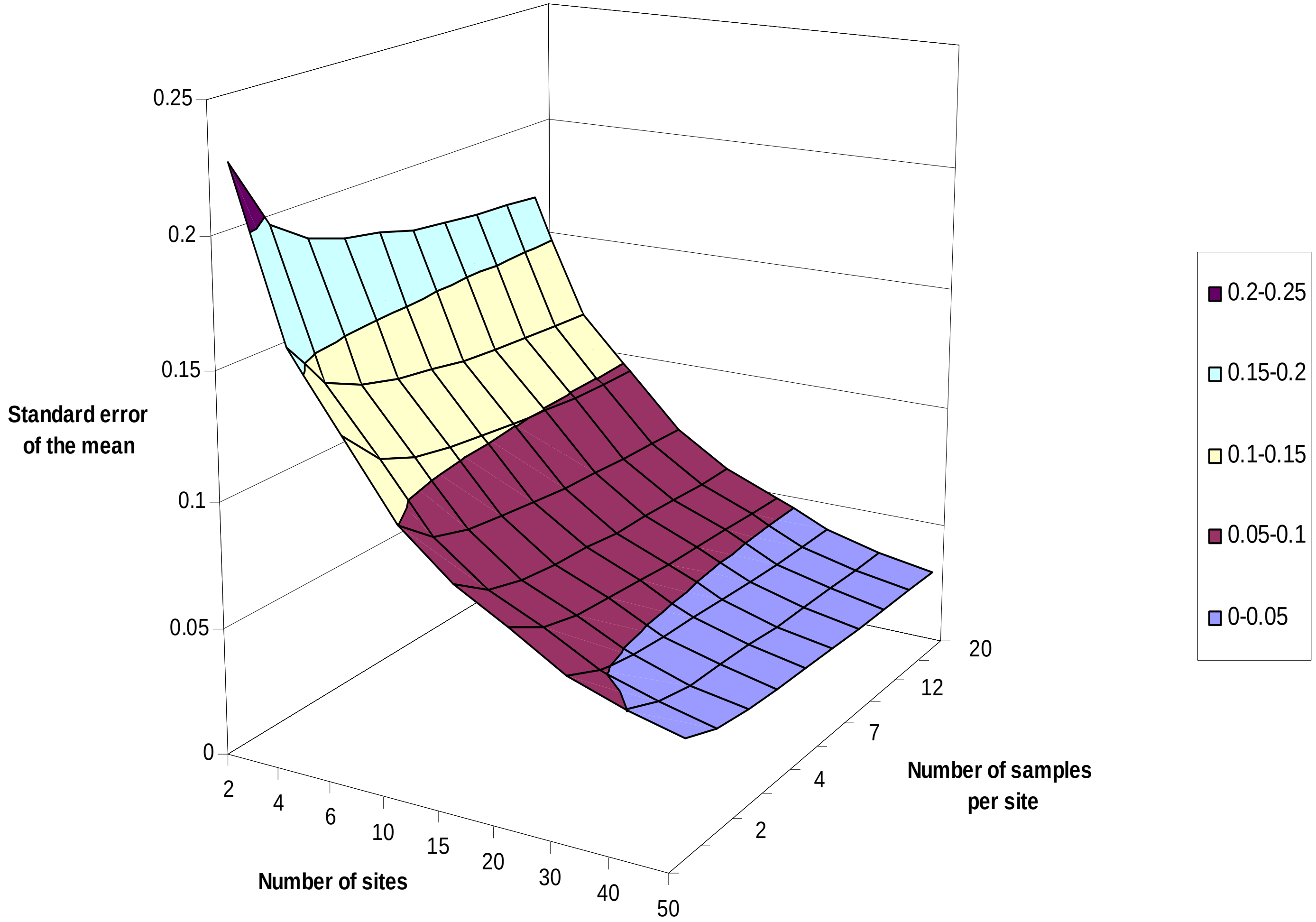


Figure B-14. Log10 Data. Standard Error of the Mean: Advantex BOD: Variance Between = 0.054: Variance Within = 0.047

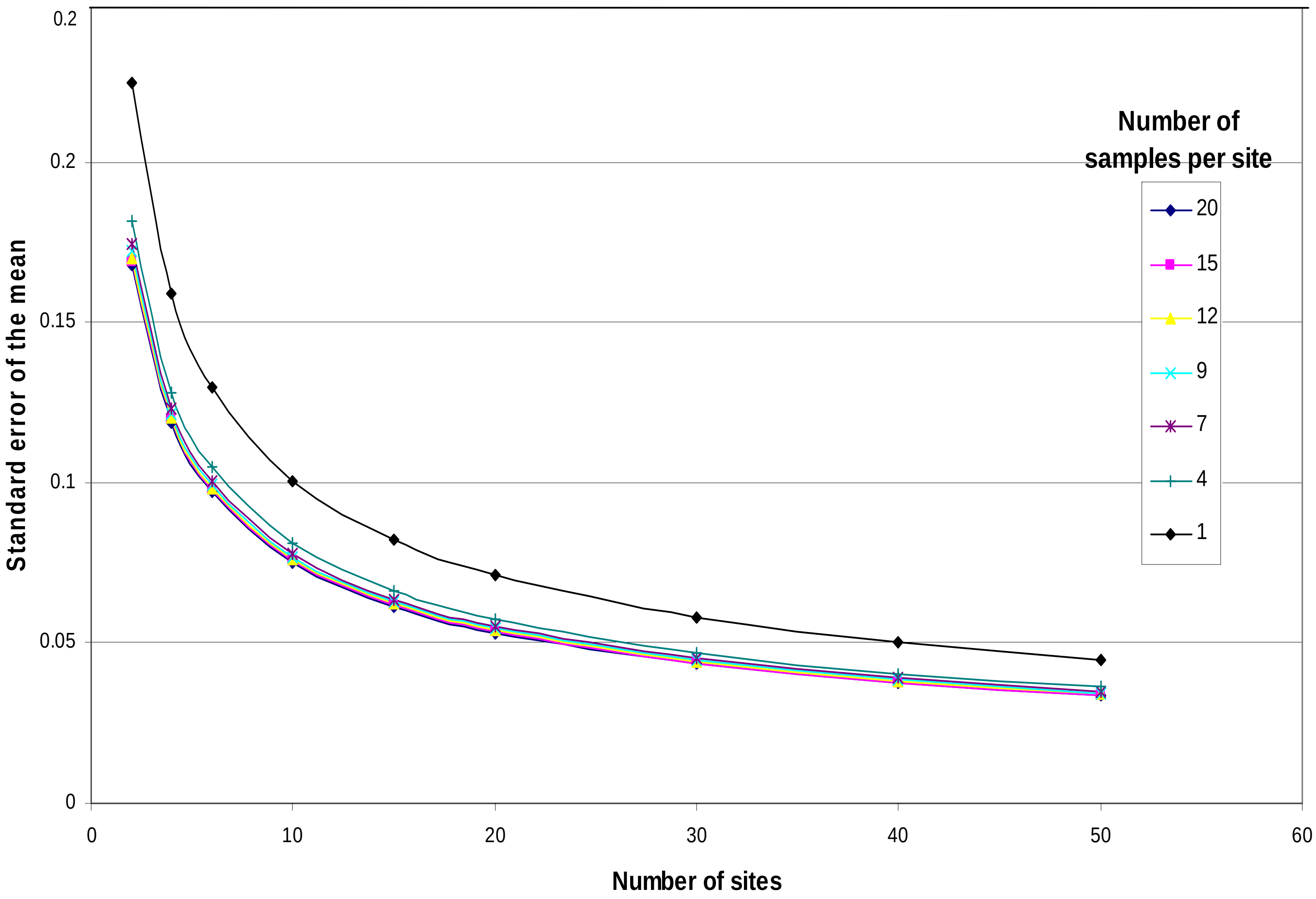


Figure B-15. Log10 Data. Standard Error of the Mean: Bioclere TSS: Variance Between = 0.093: Variance Within = 0.071

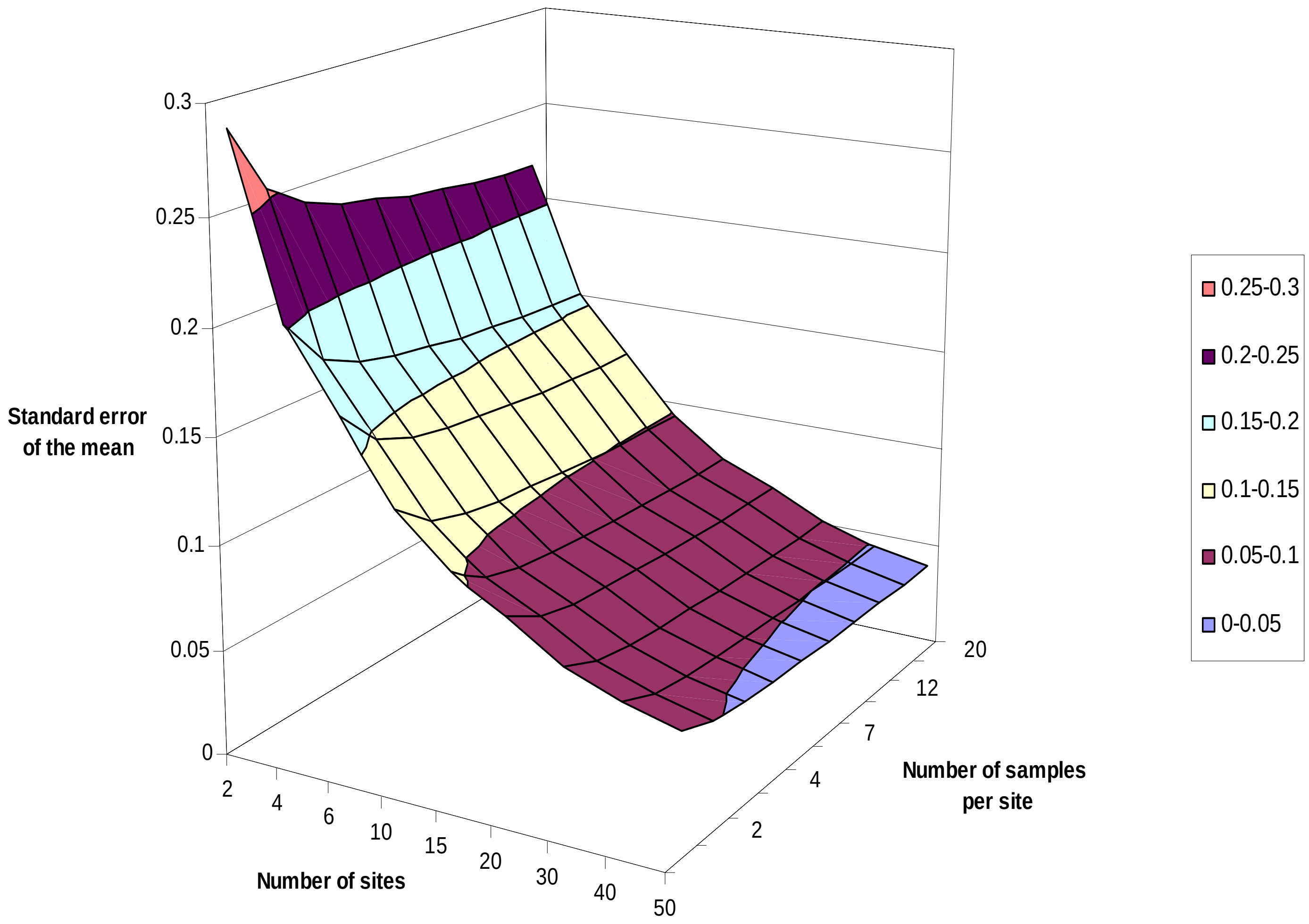


Figure B-16. Log10 Data. Standard Error of the Mean: Bioclere TSS: Variance Between = 0.093: Variance Within = 0.071

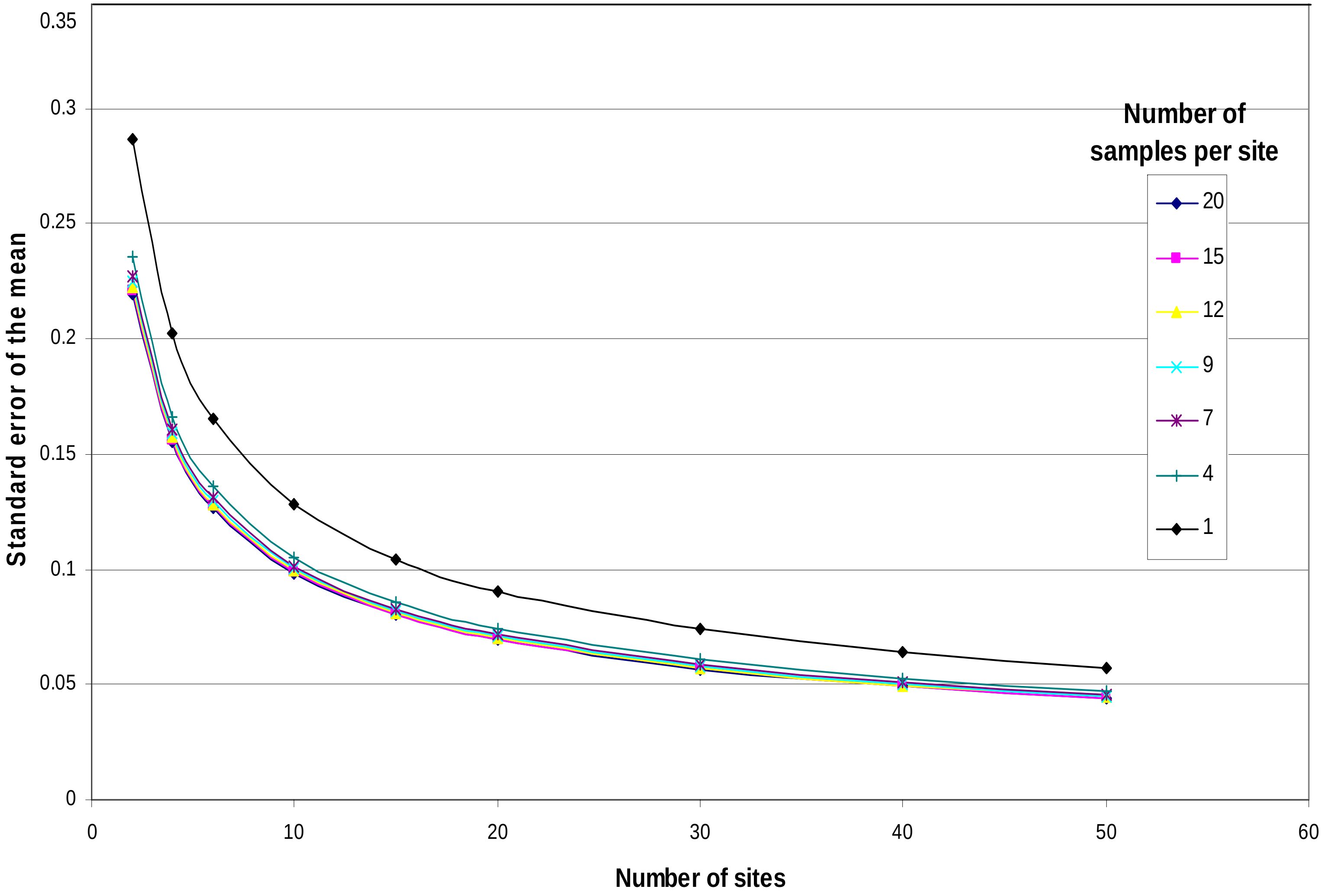


Figure B-17. Log10 Data. Standard Error of the Mean: Bioclere BOD: Variance Between = 0.089: Variance Within = 0.057

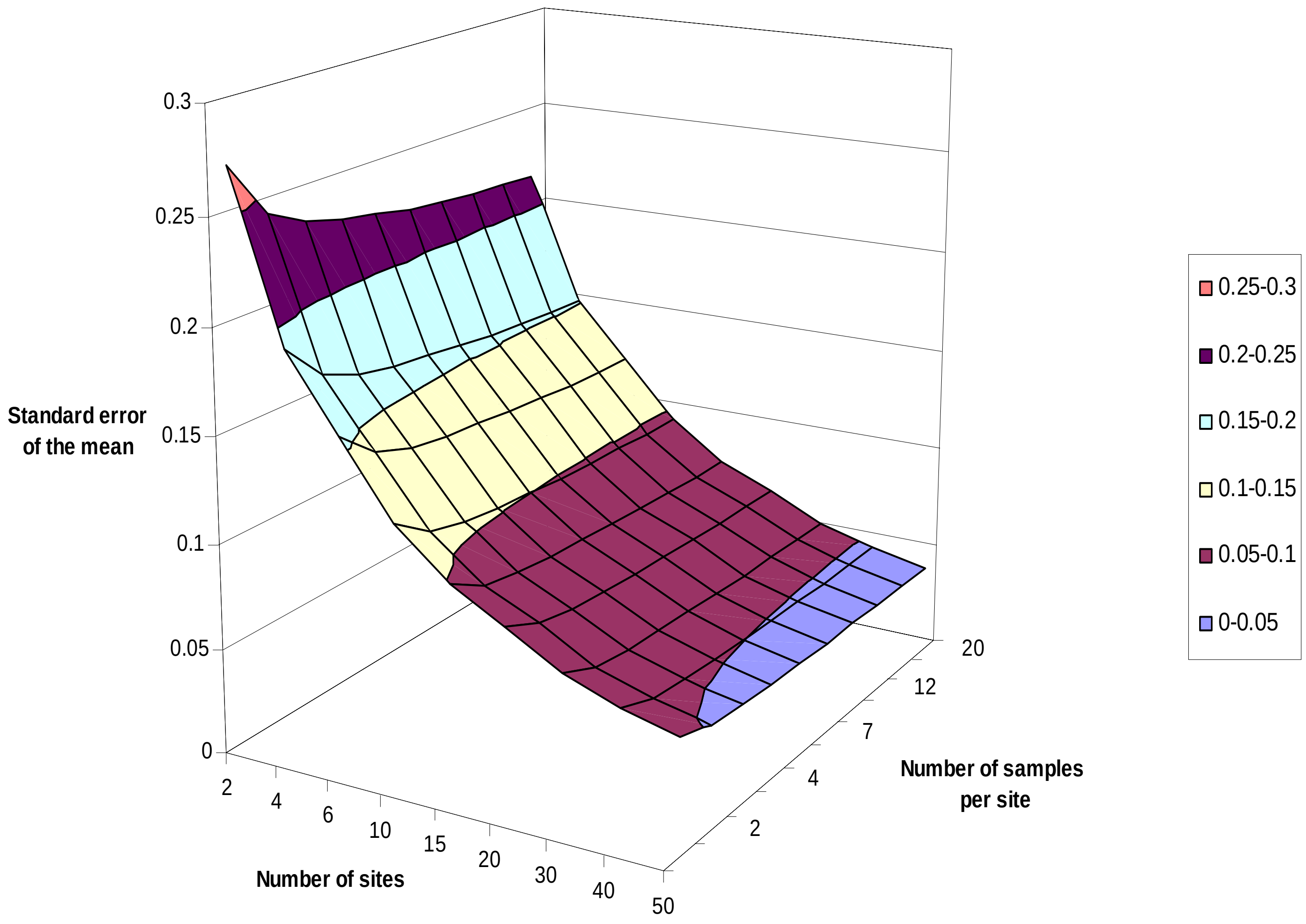


Figure B-18. Log10 Data. Standard Error of the Mean: Bioclere BOD: Variance Between = 0.089: Variance Within = 0.057

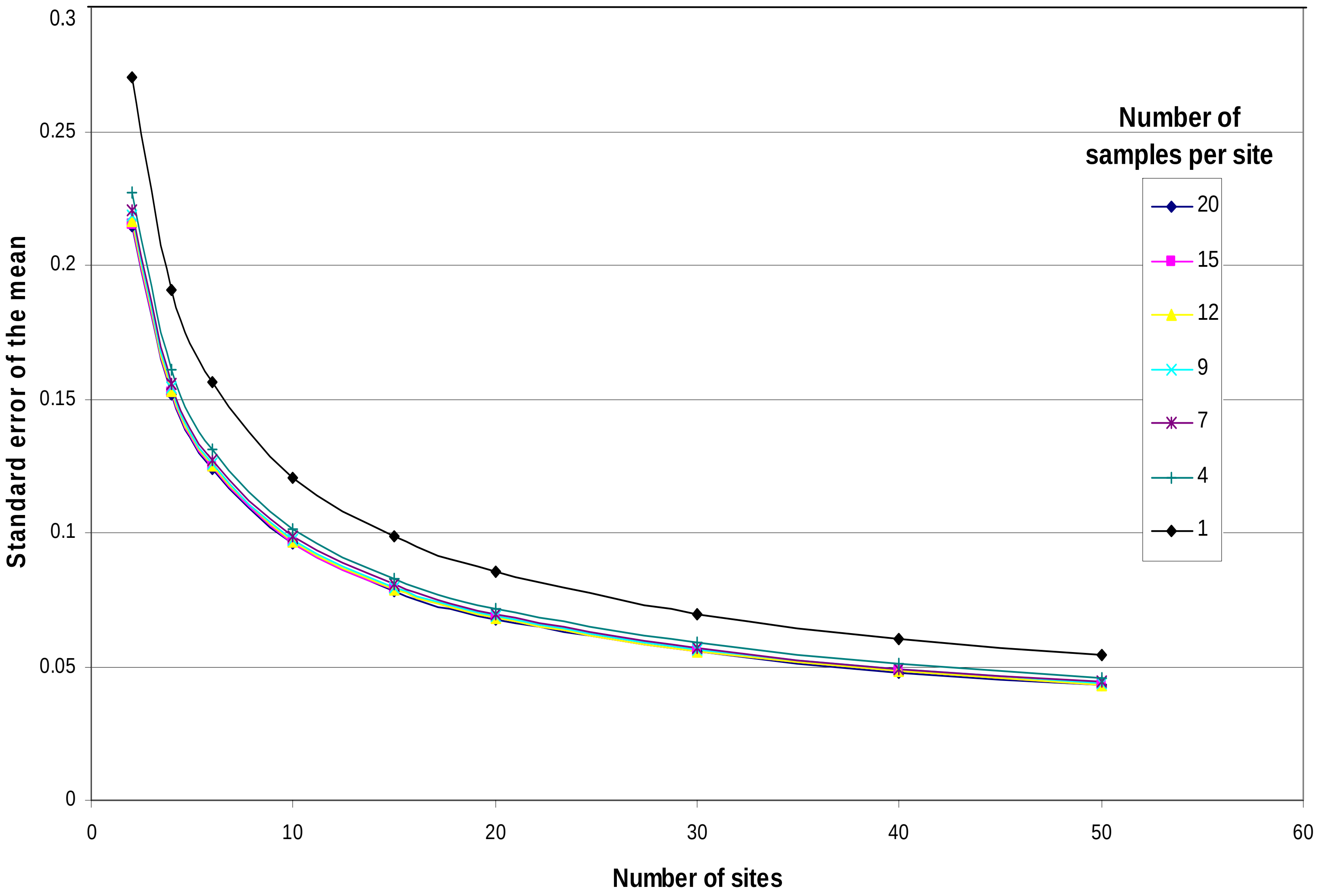


Figure B-19. Log10 Data. Standard Error of the Mean: FAST TSS: Variance Between = 0.057: Variance Within = 0.051

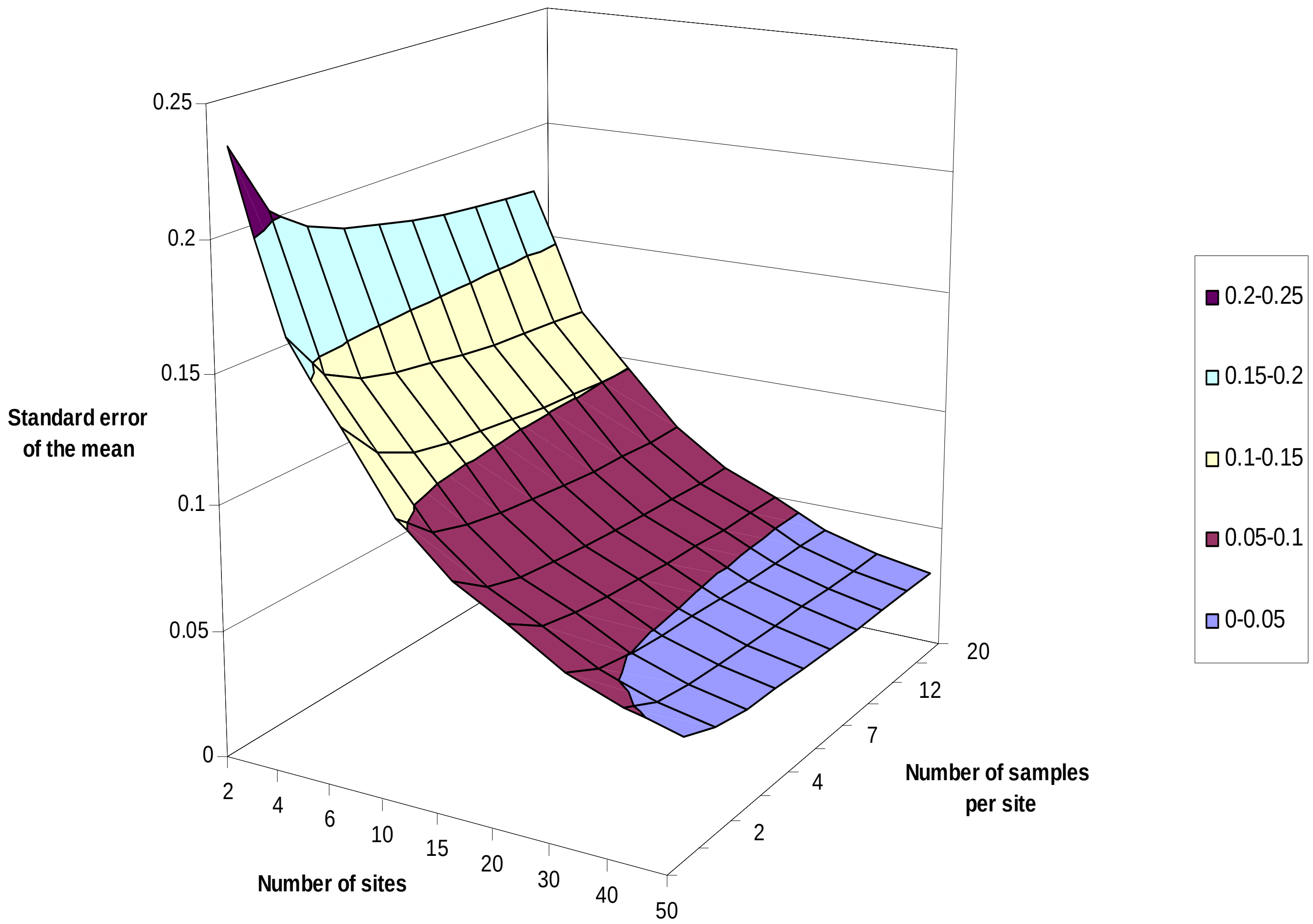


Figure B-20. Log10 Data. Standard Error of the Mean: FAST TSS: Variance Between = 0.057: Variance Within = 0.051

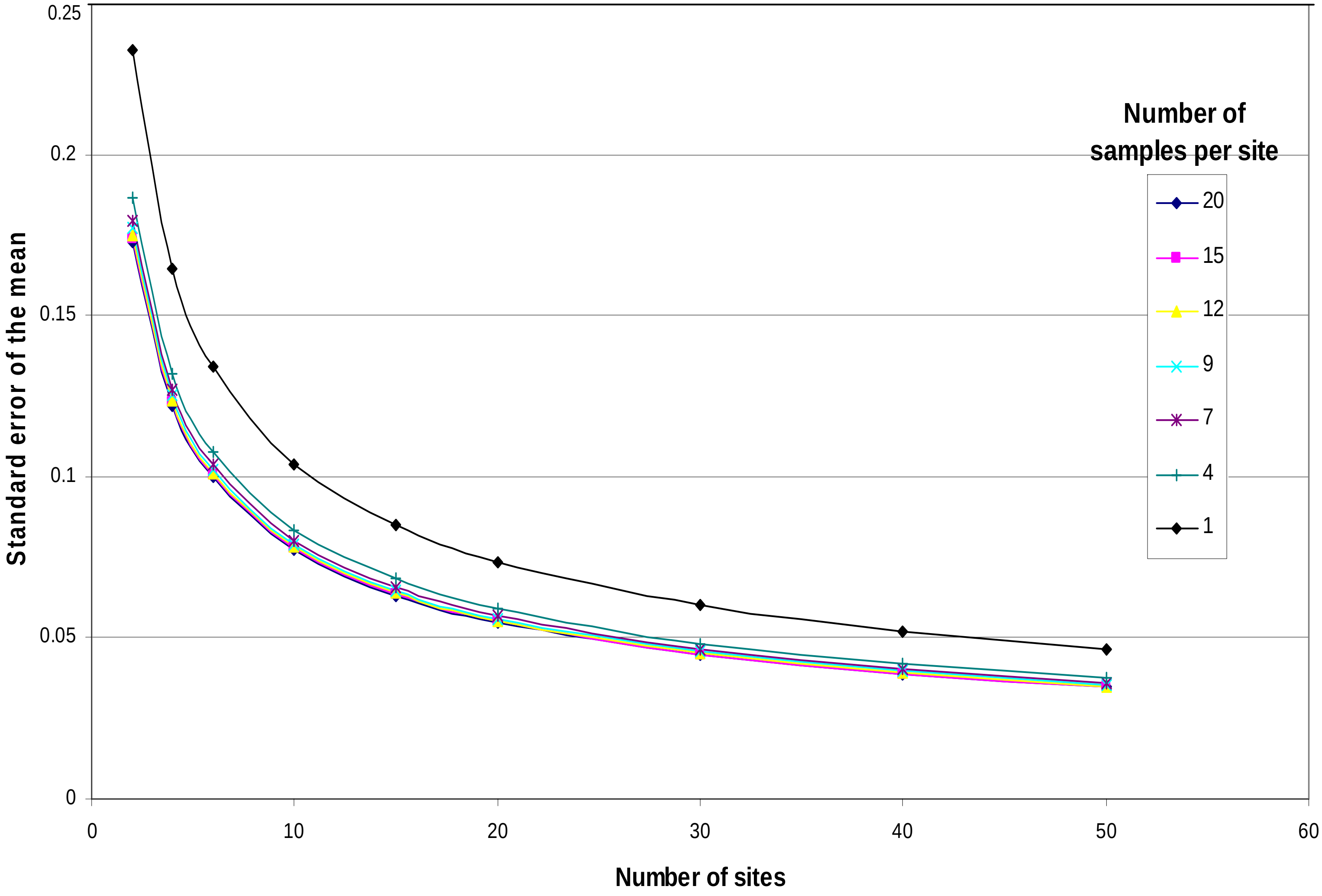


Figure B-21. Log10 Data. Standard Error of the Mean: FAST BOD: Variance Between = 0.064: Variance Within = 0.038

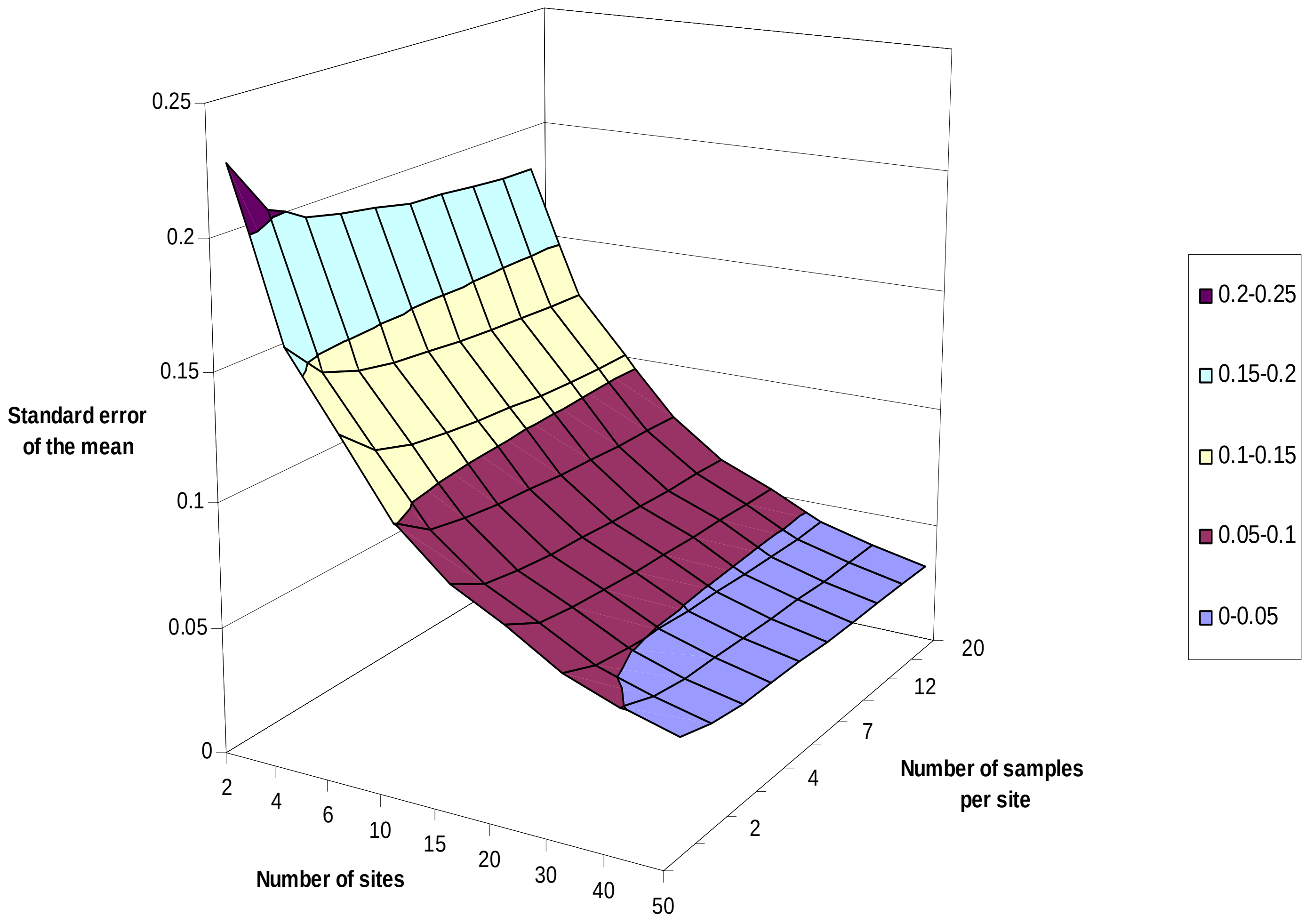
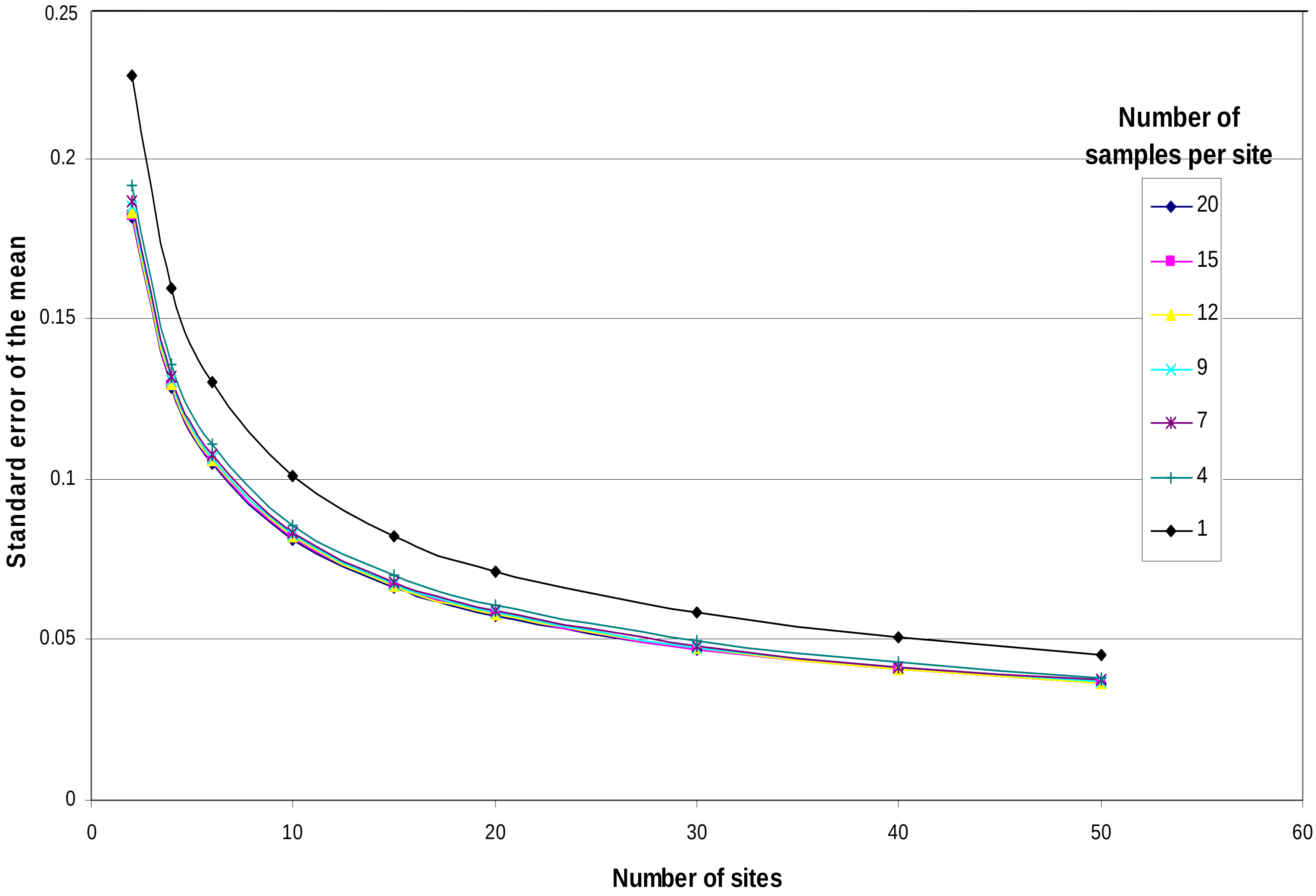


Figure B-22. Log10 Data. Standard Error of the Mean: FAST BOD: Variance Between = 0.064: Variance Within = 0.038



C QUALITY ASSURANCE PROJECT PLAN (QAPP) AND QAPP FINAL REPORT

The QAPPs are available electronically on the CD Resource Tool and online at www.ndwrcdp.org.

- Quality Assurance Project Plan
- QAPP Final Report

D DECISION SUPPORT SYSTEM (DSS): SAMPLE SPREADSHEETS, INSTRUCTIONS, AND TUTORIAL

The DSS is available electronically on the CD Resource Tool and online at www.ndwrcdp.org.

- Sample spreadsheets filled in with simulated assessments (Excel)
- Sample spreadsheets to fill out (Excel)
- Explanations and instructions in a portable document file (pdf)
- Microsoft PowerPoint overview on the use of the DSS

Printed on recycled paper in the United States of America.

WU-HT-03-35

NDWRCDP
Washington University, Campus Box 1150, One Brookings Drive, Cupples 2, Rm. 11, St. Louis, Missouri 63130-4899 • USA

This report is available online at www.ndwrcdp.org.This report is also available through the
National Small Flows Clearinghouse • West Virginia University/NRCCE, P.O. Box 6064, Morgantown,WV 26506-6064 • USA
Tel: (800) 624-8301 • WWCDRE51

NDWRCDP Variability and Reliability of Test Center and Field Data: Definition of Proven Technology From a Regulatory Viewpoint WU-HT-03-35