SHIP PRODUCTION COMMITTEE
FACILITIES AND ENVIRONMENTAL EFFECTS
SURFACE PREPARATION AND COATINGS
DESIGN/PRODUCTION INTEGRATION
HUMAN RESOURCE INNOVATION
MARINE INDUSTRY STANDARDS
WELDING
INDUSTRIAL ENGINEERING
EDUCATION AND TRAINING

August 16, 2000
NSRP 0561
N1-97-1, Subtask 33

THE NATIONAL SHIPBUILDING RESEARCH PROGRAM

Development of a Quick TBT Analytical Method

U.S. DEPARTMENT OF THE NAVY
CARDEROCK DIVISION,
NAVAL SURFACE WARFARE CENTER

in cooperation with
National Steel and Shipbuilding Company
San Diego, California

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
16 AUG 2000	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
The National Shipbuilding Research Program, Development of a Quick TBT Analytical Method	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Surface Warfare Center CD Code 2230-Design Integration Tools Bldg 192, Room 128 9500 MacArthur Blvd Bethesda, MD 20817-5700**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**SAR**	**41**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

DISCLAIMER

These reports were prepared as an account of government-sponsored work. Neither the United States, nor the United States Navy, nor any person acting on behalf of the United States Navy (A) makes any warranty or representation, expressed or implied, with respect to the accuracy, completeness or usefulness of the information contained in this report/ manual, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or (B) assumes any liabilities with respect to the use of or for damages resulting from the use of any information, apparatus, method, or process disclosed in the report. As used in the above, “Persons acting on behalf of the United States Navy” includes any employee, contractor, or subcontractor to the contractor of the United States Navy to the extent that such employee, contractor, or subcontractor to the contractor prepares, handles, or distributes, or provides access to any information pursuant to his employment or contract or subcontract to the contractor with the United States Navy. ANY POSSIBLE IMPLIED WARRANTIES OF MERCHANTABILITY AND/OR FITNESS FOR PURPOSE ARE SPECIFICALLY DISCLAIMED.

Development of a Quick TBT Analytical Method

FINAL REPORT

For the
National Shipbuilding Research (NSRP)
Environmental Science and Technology Program

by

222 East Main Street
Norfolk, VA 23506
Tel. (757) 622-2137, Fax. (757) 622-2490
E-Mail: casrm@infi.net

Dr. Thomas J. Fox, Executive Director

June 20, 2000

Table of Contents

Executive Summary

Concern about the toxic effect of tributyltin have caused the Commonwealth of Virginia through the Department of Environmental Quality to promulgate a Virginia Pollutant Discharge Elimination System (VPDES) discharge standard of 50 parts per trillion (pptr or ng/L). In addition to this action, an international ban on the use of TBT paints is scheduled for 2008. These actions as well as the interest by some shipyards in ISO 14001 certification have focused attention on removing tributyltin from shipyard discharges to the environment. One manner in which shipyards can mitigate TBT discharges from shipyard waters is through treatment of waters containing TBT and testing of waters thought to contain TBT. In order to determine whether a water contains TBT or if a water is undergoing treatment to know whether adequate treatment has been achieved, a fast turnaround method is needed. A fast analytical method is desirable since holding water at a shipyard can negatively impact shipyard production and impose a cost of containment.

To address these issues, a study was conducted to develop a "Fast-TBT Method" that would allow analysis of aqueous samples that could produce analytical results in less than half a day from the time a sample was received. Specifically for a calibrated instrument, the goal was to develop a method that would allow a single measurement of a sample in less than fifteen minutes.

The effort described in this report meets the goals established for the method. A single injection/stripping of a sample can be conducted in less than 15 minutes and using sample triplicates and standard additions for quality assurance and quality control. The samples included in this effort included high ionic strength sonar dome samples as well as ship hull wash water samples. Comparison of sonar dome samples split with a laboratory using the Virginia regulatory -approved provided initial data that the Fast-TBT Method provided comparable results to the approved method.

An inter-laboratory (round robin) comparison study was conducted as a follow-on effort to further attempt to validate the Fast-TBT Method. In this effort, twenty TBT-spiked samples were created in one laboratory, split and handled identically, and delivered to participating laboratories over the course of a four-week period. These (blind) samples ranged from less than 10 ng/L (pptr) up to approximately 10,000 ng/L (pptr). Analytical results were delivered to a third party not directly connected to the participants in the inter-laboratory comparison and statistically analyzed. The results indicate that the methods were comparable in the range at which shipyards would likely be discharging treated (or naturally low concentration) TBT-containing water and that regulatory acceptance of the method has a high probability of occurring if a more extensive effort (number of participants and a number of sample matrices) is employed.

Background: The Impact of TBT From Antifouling Paints on the Environment

Tributyl tin (TBT) has been shown to be highly toxic to certain aquatic organisms at concentrations measured in the low parts per trillion (ppt). While it is used in a wide variety of industrial and commercial products, of most concern is its use in antifouling paints applied to ship hulls.

Antifouling paints are used to coat the bottoms of ships to prevent sealift –attachment of biological organisms such as algae and mollusks to ship hulls. The attachment of marine growth to ship hulls has a negative economic impact since their presence increase hull roughness and friction, thus, increasing fuel consumption. In the early days of sailing ships, lime and later arsenic were used to coat ships' hulls, until the modern chemicals industry developed effective antifouling paints using metallic compounds.

Modern antifouling paints are typically designed to slowly leach toxic compounds into sea water, killing barnacles and other marine life that have attached to the ship. Studies have shown that these compounds can persist in the water, killing sealift, harming the environment and possibly entering the food chain. One of the most effective antifouling paints, developed in the 1960s, contains the organotin tributyltin (TBT), which has been proven to cause deformations in oysters and sex changes in whelks.

Concern about the adverse effects of TBT in the aquatic environment has been growing for many years. In 1974, oyster growers first reported the occurrence of abnormal shell growth in *Crassostrea gigas*, the pacific oyster along the East Coast of England (Key et al., 1976). However, it wasn't until the mid 1980s, that researchers in France and the United Kingdom began to suggest that the use of TBT in antifouling paints was adversely impacting a number of marine species other than the fouling organisms. This economically important species is *Crassostrea gigas*, the pacific oyster, which is farmed in coastal waters of England and France (Alzieu, 1991; Davis et al., 1988; Thain, 1983; Thain et al., 1987; Waldock, 1986; His and Robert, 1983-1985; His, 1996 and references therein). Subsequently imposex (development of penis) in female dogwelks was correlated to the presence of TBT in coastal waters (Davies and Bailey, 1991; Gibbs and Bryan, 1996; Gibbs et al., 1988, 1991; and Ten Hallers-Tjabbes et al., 1994).

However it was not until the mid 1980's that the effect of the use of TBT in antifouling paints was seen in the United States. Professor Edward D. Goldberg, of the Scripps Oceanographic Institute, has stated that "TBT is perhaps the most toxic substance ever deliberately introduced to the marine environment by mankind" (Goldberg, 1986).

During the middle and late 1980's a number of countries introduced legislation to control the use of TBT in coastal waters. Typical of these efforts was the US Federal "Organotin Paint Control Act" of 1988. Organotins are the only chemical compound regulated by law in the United States in which environmental legislation has been enacted solely for the chemical by name. The purpose of the Act was "to protect the aquatic environment by reducing immediately the quantities of organotin entering the waters of the United States." The prohibitions in the Act are:

"No person in any State may apply to a vessel that is less than 25 meters in length an antifouling paint containing organotin" with the following exceptions: "(1) the aluminum hull of a vessel that is less that 25 meters in length; and (2) the outboard motor or lower drive unit of a vessel that is less than 25 meters in length." No person in any State may: (1) sell or deliver to, or purchase or receive from, another person an antifouling paint containing organotin; or (2) apply to a vessel an antifouling paint containing organotin; unless the antifouling paint is certified by the Administrator [of EPA] as being a qualified antifouling paint containing organotin, and (3) sell or deliver to, or purchase or receive from, another person at retail any substance containing organotin for the purpose of adding such substance to paint to create an antifouling paint. A key certification was that the EPA Administrator shall certify each antifouling paint containing organotin that the Administrator has determined has a release rate of not more than 4.0 micrograms per square centimeter per day.

The years since the passage of the Organotin Paint Control Act in the US and other similar laws in other countries around the world have resulted in a general reduction of TBT levels in many former “hot spots” such as marinas and harbors. However, concern over TBT remains. This concern has resulted in international action through the International Maritime Organization (IMO) to ban the use of TBT in antifouling paints worldwide.

The harmful environmental effects of organotin compounds were recognized by IMO in 1990, when the Marine Environmental Protection Committee (MEPC) of the IMO adopted a resolution which recommended that Governments adopt measures to eliminate the use of antifouling paint containing TBT on non-aluminum hulled vessels of less than 25 meters in length and eliminate the use of antifouling paints with a leaching rate of more than 4 microgram’s of TBT per day. Some countries, such as Japan, have already banned TBT in antifouling paint for most ships.

The 21st IMO Assembly, held in London in November 1999, passed a resolution which includes a proposed deadline of 2008 for the complete prohibition of organotins acting as biocides in antifouling systems on ships. The IMO ban calls for a halt to application of TBT antifouling paints starting 2003, with a complete ban on the use of TBT paint on ships worldwide by 2008. It is anticipated that during the years 2003 to 2008 there will be a surge in ship-repair work required to repaint approximately 30,000 oceangoing ships.

Impact of TBT Discharge Limits on Shipyard Operations in Virginia

The Commonwealth of Virginia, uniquely among U.S. states, has chosen to regulate the discharge of TBT from Virginia shipyards under the National Pollution Discharge Elimination System (NPDES). By summer 2000, all Virginia shipyards may discharge wastewater containing no more than 50 parts per trillion (ppt), of TBT to the surface waters of the State. To complicate matters even further, the regional wastewater treatment authority will not accept water with any level of TBT.

Shipyards generate large quantities of TBT in wash water used to wash down the hulls of ships coated with TBT. The wash down of a panamax size ship can generate over 100,000 gallons of water containing up to 1,000,000 ppt TBT.

At the time the Virginia limit was imposed, there was no technology available which shipyards could use to remove TBT from wastewater to meet the 50 ppt permit limit. In order to achieve compliance, Virginia shipyards began a cooperative program to develop and test technology to remove TBT from shipyard wash waters . This was done through the Center for Advanced Ship Repair and Maintenance (CASRM), a non-profit corporation established as a partnership between the ship repair industry and Old Dominion University. The US Environmental Protection Agency (EPA), the Virginia Department of Environmental Quality, and the industry have funded elements of this project.

The Necessity for a Rapid Method for TBT Analysis.

When this project was started it was soon determined that a key requirement for success would be the development of a rapid method to analyze water samples for TBT at levels of 50 ppt or below. The industry standard method used by many commercial and government laboratories has a detection limit of 1 ppt, but because the analysis method is complicated and labor intensive, each sample may take two days or more to complete. Development of a rapid TBT analytical method became a high priority since it would allow the research team to make multiple TBT measurements needed to optimize the wash water treatment process. In addition a quick turnaround method would allow monitoring of the treatment process at shipyards and could provide feedback during a treatment run that could allow changing treatment parameters to meet treatment goals.

To meet this need, the Environmental Science and Technology program of the NSRP SP-1 (environmental) panel sponsored the development of the new method. A description of method development and the analytical method are provided below.

Description of the Rapid TBT Analytical Method

As a highly toxic biocide whose release into natural waters poses a risk for many aquatic species, a variety of analytical methods have been developed to determine tributyl tin (TBT), and its immediate degradation products, dibutyl tin (DBT) and monobutyl tin (MBT). These methods fall into two classes: “Direct”, ones which use sodium borohydride to convert the dissolved butyl tin cations to their corresponding hydrides (TBTH, $DBTH_2$, and $MBTH_3$) with subsequent collection using cryotrapping and modified atomic absorption spectrometry (AAS; e.g., Hodge et al., 1979; Valkirs et al., 1987); and “Indirect”, ones using either hydride formation followed by solid phase collection, solvent extraction, and GC/FPD (gas chromatography/flame photometric detection; e.g., Sharron et al., 1995) or immediate solvent extraction and GC/FPD (e.g., Matthias et al., 1986), or solvent extraction, alkylation with a Grignard reagent, and detection with GC/FPD or GC/mass spectrometry (e.g., Unger et al., 1986). Each of these methods has their advantages and disadvantages, but for the analysis of industrial process waters (e.g., from

shipyard operations), a direct method has the advantage of being relatively fast to allow the analysis of up to 10 samples per day, and having sufficiently low detection limits (ca., 10 ng Sn/L as TBT) to meet stringent monitoring requirements. Therefore, we adapted the hydride generation/AAS method of Hodge et al. (1979) to allow the rapid analysis of untreated and treated waters from shipyard operations.

Apparatus. The system used to generate, collect, and quantify the butyl tin hydrides are similar to those described in Hodge et al. (1979) with several exceptions. The glass stripping vessel is slightly larger (29 cm x 3.9 cm, with 34/45 ground glass joint), holding a sample volume of up to 150 mL, while a water trap consisting of a 20 cm x 1/4" (3/16" ID) FEP tubing packed with 4-8 mesh anhydrous calcium chloride (dimethyl dichloro silane (DMCS)-treated glass wool at each end) is placed between the stripper and liquid nitrogen trap. The liquid nitrogen trap is an 18 cm x 1/4" borosilicate glass tube bent into a V-shape, packed with DMCS-treated glass wool, and having 1/4"-1/8" stainless steel Swagelok reducing unions (equipped with Teflon ferrules) at each end; this entire trap and fittings assembly is wrapped with Ni-Cr wire attached to a Variac transformer to heat it to ca. 180 ^{o}C. The quartz tube burner is identical to that of Hodge et al. (1979), all tubing and fittings are Teflon (except the liquid nitrogen trap), and all glassware is deactivated with DMCS. The burner is placed in a Buck Scientific 210 VGP Atomic Absorption Spectrometer without background corrector and equipped with a Buck hollow cathode Sn lamp; a wavelength of 286.3 nm is employed. The analog signal from the instrument is processed using a Peak Simple A/D converter and chromatographic integrator software in a notebook computer. Gas flow rates are as follows: He stripping/carrier gas, 140 mL/min; burner gases, H_2, 330 mL/min, and air, 200 mL/min.

Reagents. A 2M hydrochloric acid solution is made from Baker “Instra-analyzed” hydrochloric acid, a 4% (w/v) aqueous borohydride solution is made daily using Alfa sodium tetrahydridoborate without the addition of sodium hydroxide for stabilization, and all water used for reagents or sample dilutions is >18 megohm-cm deionized water. High purity (Alfa) $MBTCl_3$, $DBTCl_2$, TBTCl are used to make gravimetric stock 1000 ppm Sn standards in spectrometric grade methanol, and 1 ppm Sn working standards are made daily from this stock standard by dilution with deionized water.

Method. The sample volume is placed in the bottom of the stripper and ranges from 0.1 to 150 mL depending on sample concentration, and the final volume in the stripper is adjusted to 150 mL using deionized water. To this solution, 1.5 mL of 2M HCl is added and the upper and lower portions of the stripper are connected to begin He purging. After one minute of purging, the hydride trap is placed in liquid nitrogen and 2 mL of borohydride solution is added using a glass syringe with stainless steel needle over a one minute period. After six minutes (check that the flame is still lit), the trap is removed from the liquid nitrogen and the integrator software activated. After the $DBTH_2$ peak elutes (ca. 2.6 min), the trap is warmed using the Variac to elute the TBTH peak. Upon completion of the chromatogram (ca. 3.6 min) the stripper can be disassembled, rinsed, and a new sample introduced. The water trap should be replaced after every 6 samples. Each sample is analyzed in triplicate, blanks (triplicate) consist of deionized water analyzed as a sample, and calibration is made at least once per analytical run using five

standard additions (0-60 ng Sn as TBT depending on sample concentration) to a representative sample (i.e., perform standard additions to one sample of a group with similar matrices). The concentration of sample in ng Sn as TBT/L can be converted to ng TBT/L by dividing by 0.4093.

Results and Discussion

The early work of Hodge et al. (1979) describes the reagent concentrations and reaction times required to quantitatively convert dissolved organic tin species to their volatile hydrides. However, it was necessary to adjust these somewhat to optimize the chromatographic separation and quality of the peak shapes, and thus reduce the detection limits. In this respect, a higher He stripping/carrier flow rate, using hydrochloric acid in place of acetic acid, and heating the trap with Ni-Cr wire to elute TBTH greatly improved the peak shapes of the butyl tin hydrides (see Figures 1 and 2 at low and high TBT concentrations, respectively). Moreover, the hydride trap used by Hodge et al. (1979) had to be replaced after every 5 samples, and our addition of the water trap eliminates this time consuming task (i.e., difficult to replace since the trap is wrapped with Ni-Cr wire and connected to the Variac) and reduces tailing of the TBTH peak. Although the water trap requires replacement after every 6 samples, this only takes approximately 30 seconds using previously prepared traps; the use of a larger trap caused significant (>20%) losses of TBTH, and only the small one recommended here can be employed.

The analytical figures of merit were determined using distilled water, trace metal-clean water from the Sargasso Sea, and waters from a TBT treatment project (e.g., wash water, plant effluent). Figure 3 shows the highly linear behavior of TBT peak area as a function of varying amounts of TBT (as ng Sn) in distilled water; assuming a sample volume of 150 mL, Fig. 3 demonstrates linear response to a concentration of 610 ng Sn/L (or, 1482 ng TBT/L); using smaller sample sizes (e.g., 0.1 mL) this linear range can be extended to over 900 g Sn/L. Highly linear responses were also found in Sargasso Sea water (Figure 4) and shipyard waters (an example in Figure 5) for all of the butyl tin species. With the exception of MBT in the effluent sample (Fig. 5), the slopes for the butyl tin species are all very similar (i.e., as expected, the AAS is only responding to Sn). More significantly, the slopes of the TBT response in distilled water (Fig. 3), sea water (Fig. 4), and treatment plant effluent (Fig. 5) are statistically identical, demonstrating no analytical interference for the samples analyzed to date and suggesting high accuracy (i.e., calibration via standard additions is used to assure accuracy).

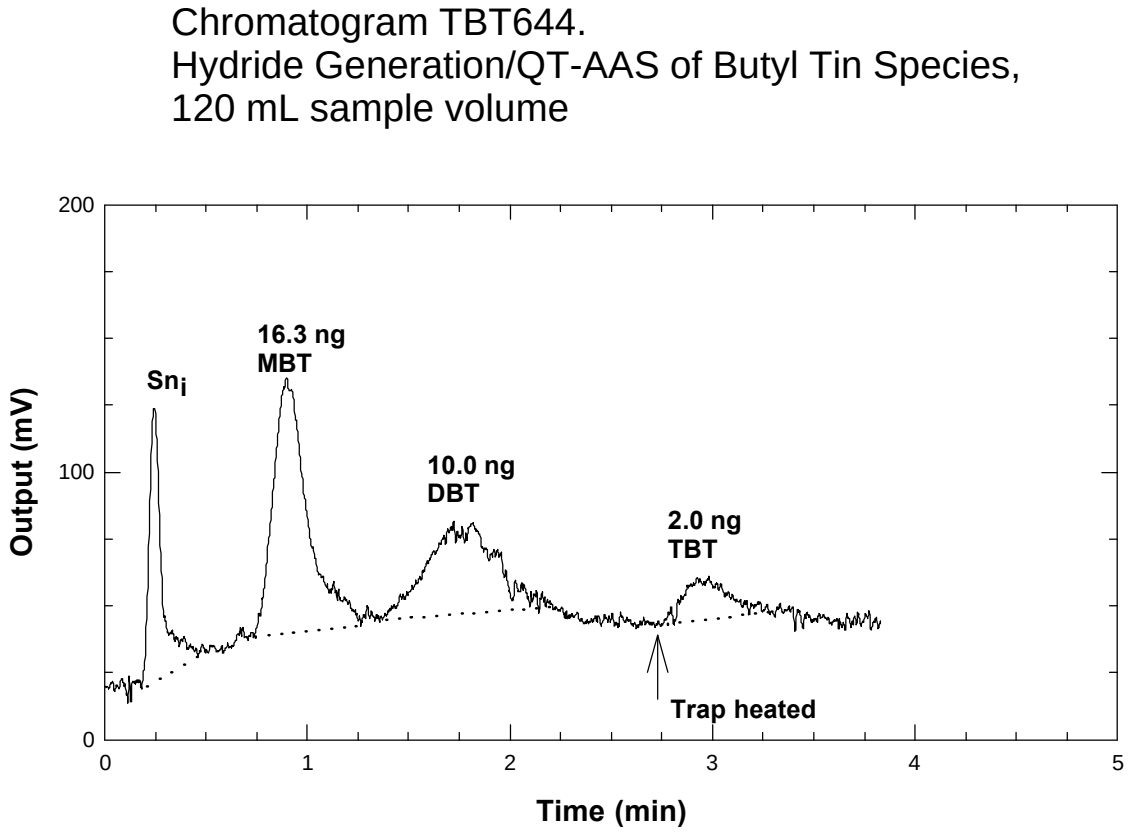

Figure 1. Example chromatogram for low TBT concentrations in shipyard samples analyzed by FAST TBT Method (17 ng TBT/L).

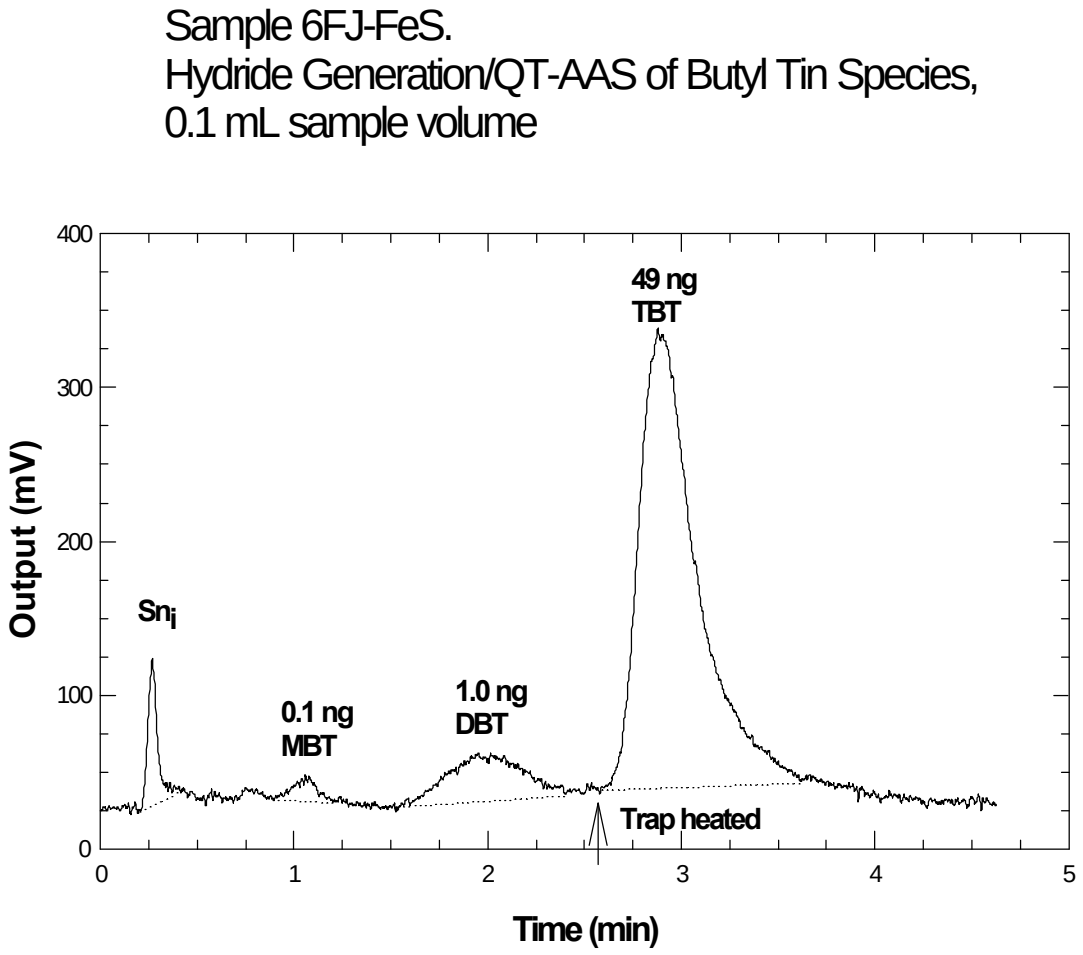

Figure 2. Example chromatogram at high TBT concentrations (490 µg TBT/L).

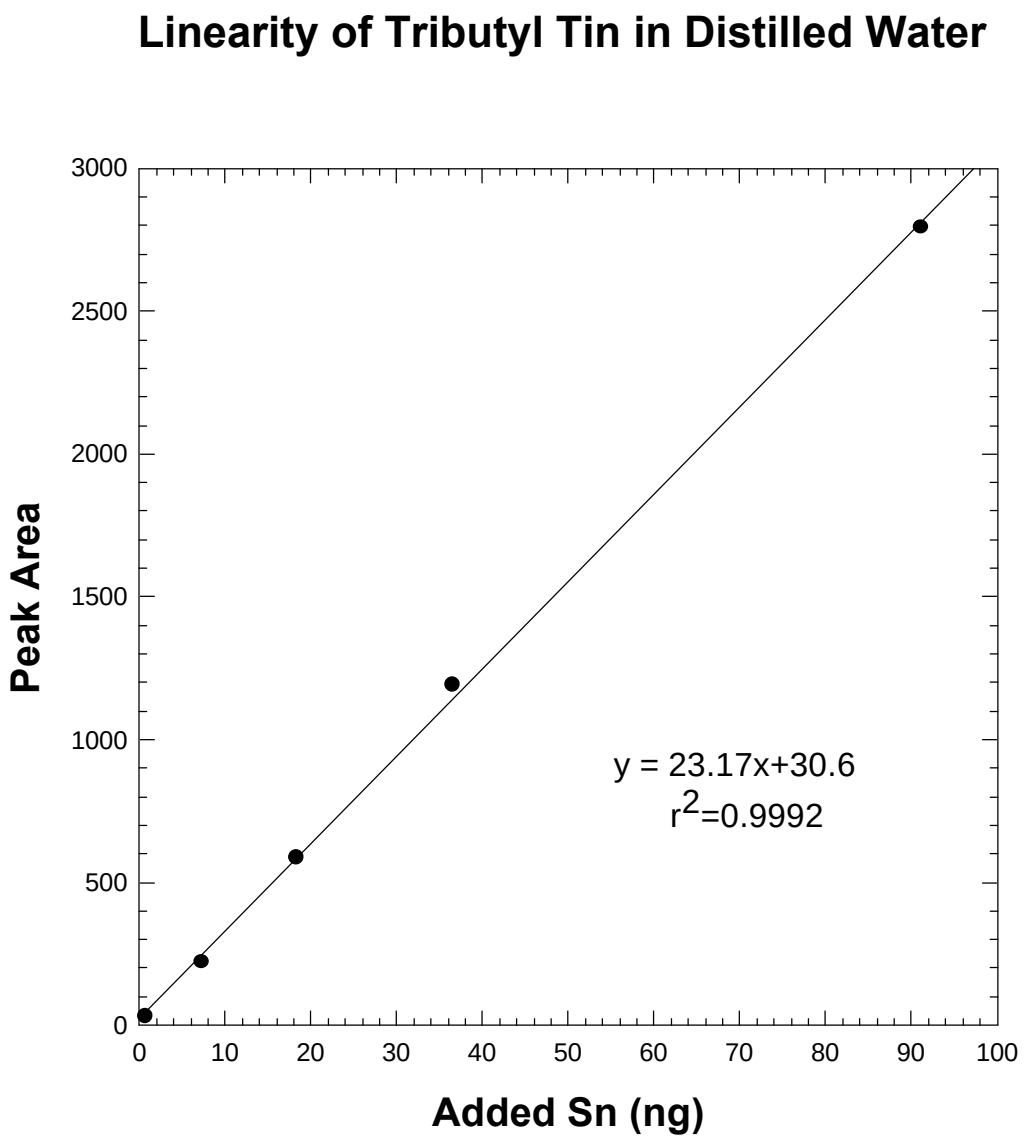

Figure 3. Example of linearity of measured response (area counts) for tributyltin in deionized distilled water.

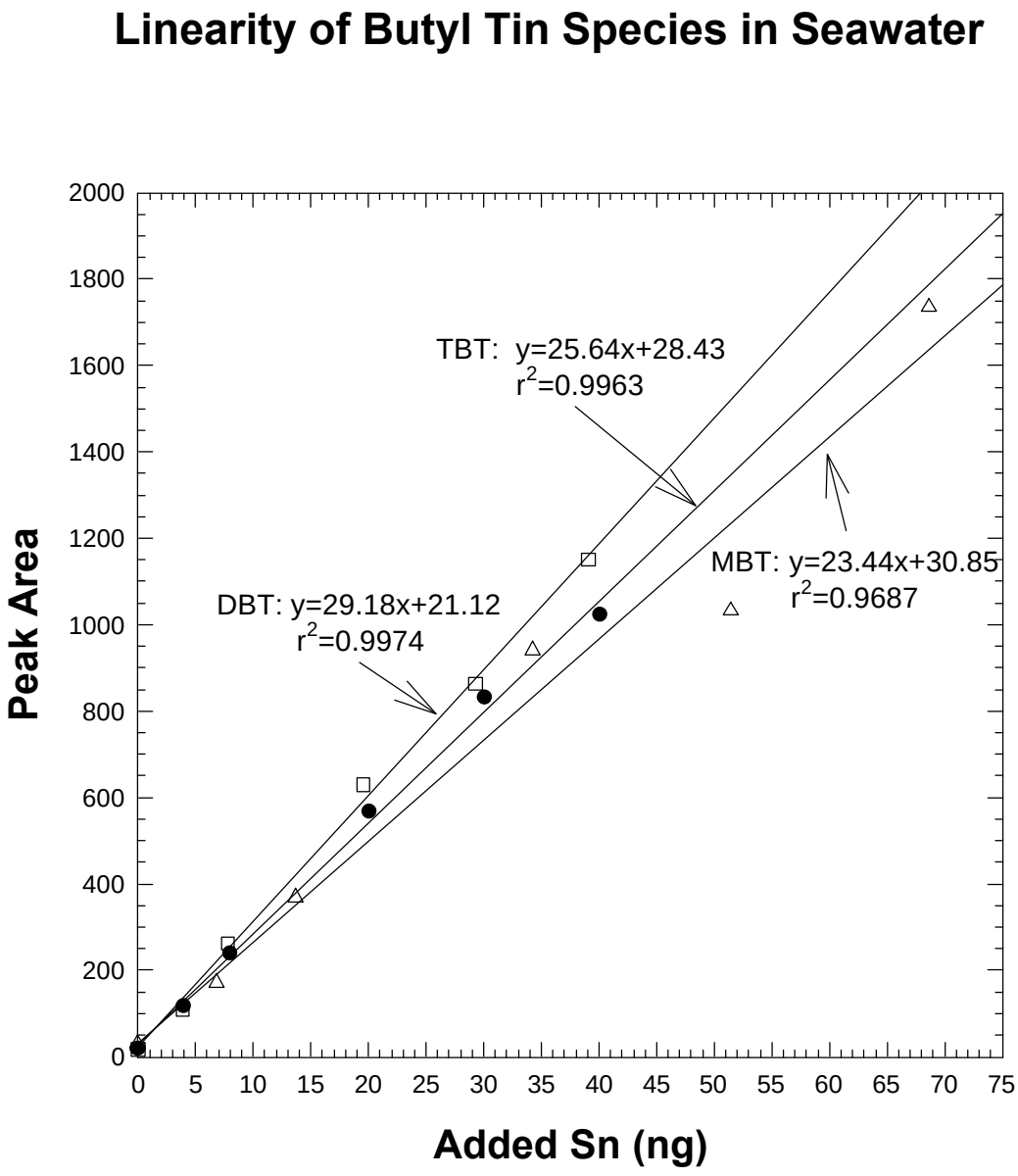

Figure 4. Measured responses of organotin species (MBT, DBT, TBT) spiked in Sargasso Seawater.

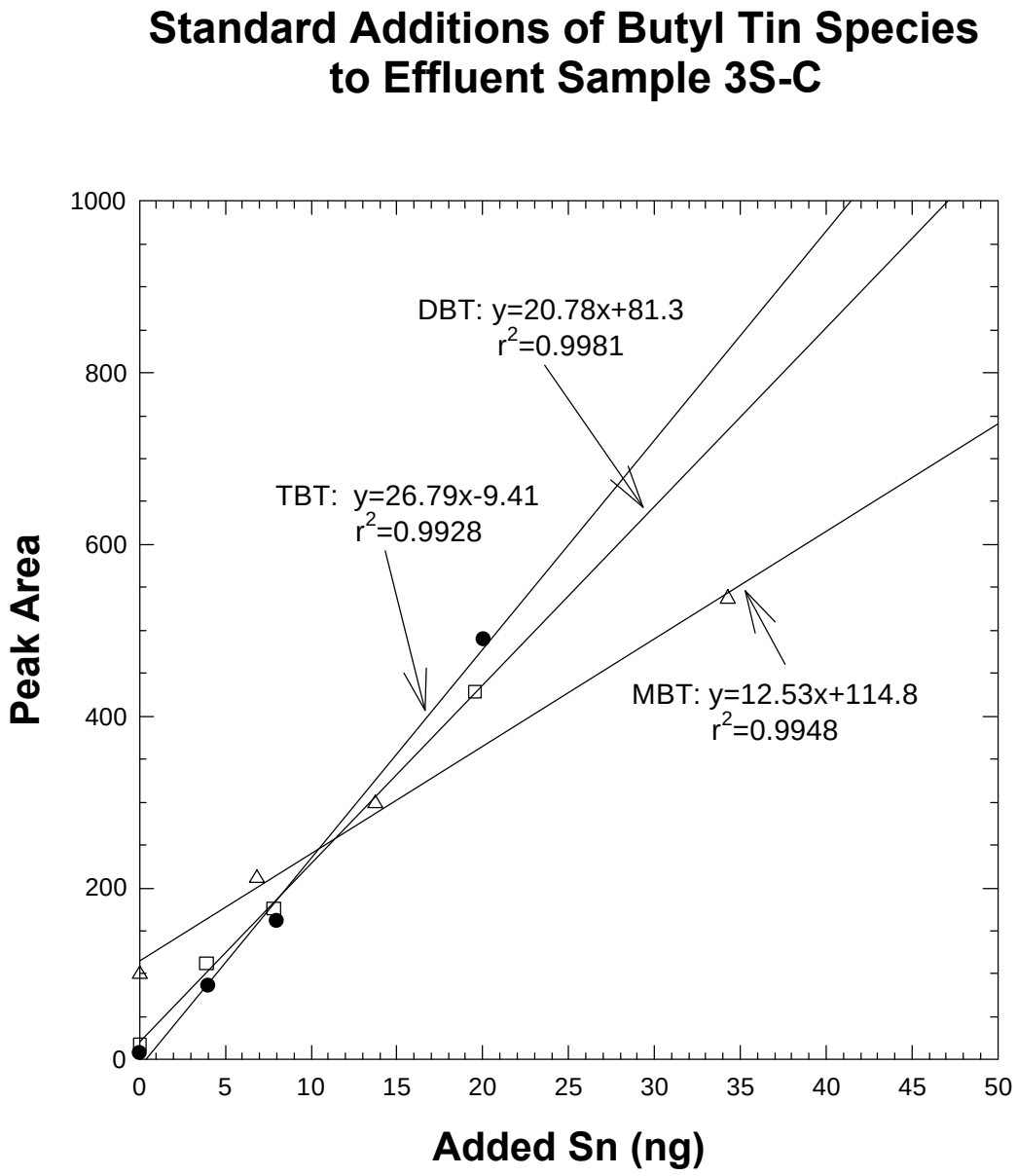

Figure 5. Hydride generation-QT/AAS responses to Butyltin species in a treated effluent sample.

Since no butyl tin has been detected in the blank with the reagents used in this method, the detection limits were determined using eight determinations of a low (0.8 ng Sn as TBT) standard (detection limit = 3 x standard deviation of a blank or low concentration). In this manner, the absolute detection limit is 0.31 ng Sn (3 ; n=8), or 0.77 ng TBT, and using a 150 mL sample volume gives a relative detection limit of 2.1 ng Sn/L (5.1 ng TBT/L). This absolute detection limit is a factor of 3 better than that given in Hodge et al. (1979), and the relative detection limit is a factor of 4 lower. The use of triplicate analyses of samples allows the precision to be evaluated directly for each sample, but work to date shows the precision (expressed as relative standard deviation) to be 7.7% (RSD) for concentrations from 5-130 ng Sn/L (12-318 ng TBT/L). The sample to sample analysis time is 11 minutes using the methods given above.

Limited Comparison of the Fast TBT Method to a Standard TBT Procedure Using Shipyard Samples

In a separately funded project, CASRM has constructed and now operates a pilot plant water treatment system to remove TBT from shipyard wastewater. It has been mounted on a barge to allow it to be moved and operated at every dry dock in the Hampton Roads, Virginia region.

The TBT treatment plant consists of a series of connected unit processes comprising a dissolved air flotation unit, a sand filter and two activated carbon columns. The treatment plant process schematic and flow diagram**.** are shown in Figure 6.

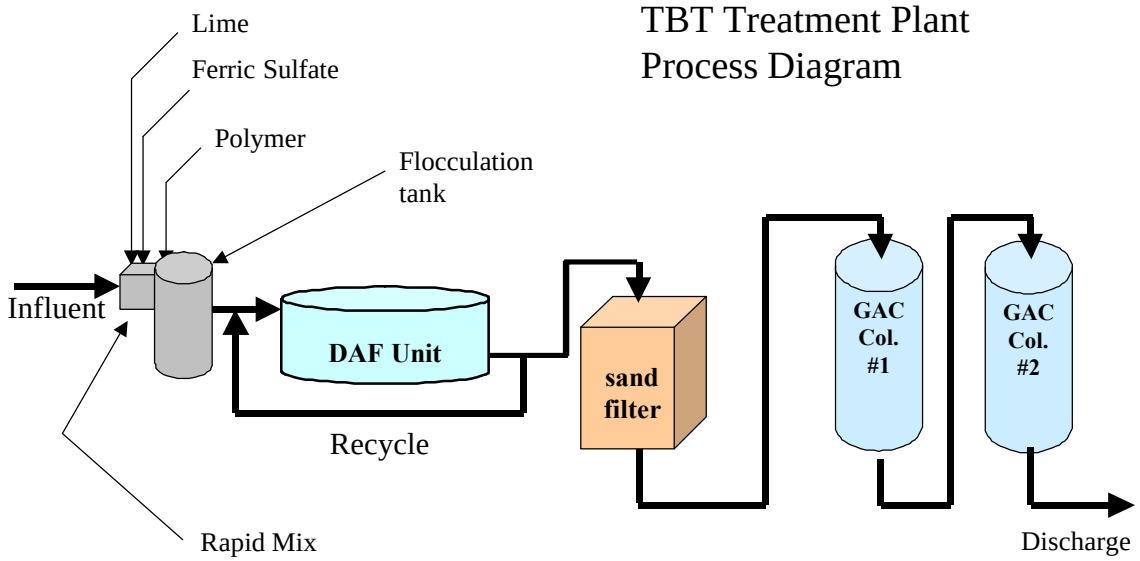

Figure 6. Treatment Process Schematic and Flow Diagram for Barge-mounted TBT Treatment Plant.

The treatment plant has been used to treat both wash water and sonar dome water that is generated by certain US Navy ships. Treatment of a sonar dome water collected from a U.S.

Navy surface combat ship while it was in dry-dock in Norfolk, VA was conducted in December 1999 and used to compare the Fast TBT method versus the accepted method in Virginia (i.e. Grignard reagent method).

Approximately 20,000 gallons of sonar dome water collected from the ship were pumped into the TBT treatment plant on December 8, 1999. Six samples were collected during the treatment plant operation. These samples included a treatment system blank (with respect to TBT), influent sample, and effluent samples from the first (GAC1) and second-stage (GAC2) activated carbon contactor columns. Samples collected during the treatment of sonar dome water were analyzed for tributyl tin by four different laboratories. All laboratories were delivered unfiltered, unpreserved (not acidified) samples in polycarbonate sample bottles delivered on ice or cold packs. The results of the analysis are included in Table 2. Results from laboratory number four were generated using the Fast TBT Method. Laboratory number one is the laboratory utilized by the Commonwealth of Virginia to analyze TBT samples for the State and has been conducting TBT analyses for over 12 years.

Table 2.
TBT Concentrations for the Sonar Dome Treatment Event: December 8, 1999.

Sample ID	Description	Lab 1	Lab 2	Lab 3	Lab 4
1S	Treatment System Blank	nd	39	23	nd
2S	Influent-Sonar Dome Water	3700	3167	2240	4083
3S	Effluent of 1st GAC Cell	17	49	45	6.6
4S	Effluent of 2nd GAC Cell	6	49	70	nd
5S	Effluent of 1st GAC Cell	21	51	26	12.9
6S	Effluent of 2nd GAC Cell	6	44	36	nd
7S	Laboratory-generated Blank	nd	44	40	nd

The results for all of the organotins measured by the Fast TBT Method for the sonar dome treatment event are presented in Table 3. Also included are results for wash water samples collected during the wash down of a commercial cruise ship. Note, the ± indicates the standard deviation for three measurements for each sample and the last column represents TBT in terms of the mass concentration of the TBT cation (all other values are in terms of the mass concentration of tin). The last five samples in the table are the results for the wash water from the cruise ship. The filtered and unfiltered concentrations of TBT exceeded 1 million ng/L (pptr) while concentrations after coagulation were lowered to between 13,000 and 200,000 ng/L. The method has been observed to be capable of working well with both low and high concentration samples, as long as only a small volume of high concentration sample is used for analysis.

Table 3
Butyl Tin Species in Shipyard and Process Samples
All determinations in triplicate

Sample ID	MBT, ng Sn/L	DBT, ng Sn/L	TBT, ng Sn/L	TBT, ng TBT/L
1S-C	<16.5	<10.4	<2.1	<5.7
2S-C (acidified)	2143 ± 106	4167 ± 214	1671 ± 151	4083 ± 369
3S-C	79.2 ± 6.1	8.61 ± 0.65	2.69 ± 0.20	6.57 ± 0.49
4S-C	13.8 ± 2.6	<10.4	<2.1	<5.7
5S-C	80.8 ± 9.6	10.5 ± 0.5	5.28 ± 0.07	12.9 ± 0.2
6S-C	NA	<10.4	<2.1	<5.7
5FJ-FeS	6369 ± 651	33781 ± 2277	105212 ± 7570	257054 ± 18496
6FJ-FeS	5460 ± 262	38323 ± 2683	199670 ± 13777	487833 ± 33660
9FJ-AlS	805 ± 105	9802 ± 996	13317 ± 263	32536 ± 643
102099F (filt.)	16251 ± 726	137428 ± 11089	448968 ± 13976	1096917 ± 34146
102099F (unfilt.)	33920 ± 2539	123706 ± 2800	471827 ± 46159	1152766 ± 112775

NA – not analyzed due to chromatographic interference

Inter-laboratory Comparison (Round Robin) With the Fast TBT Method Summary

Background:

As noted earlier in this report, in the Commonwealth of Virginia there currently exists only one accepted analytical method for the measurement of tributyltin. This method is a Grignard reagent derivitization method with solvent extraction that in Virginia is commonly referred to as the "Unger" or "VIMS" method. This method is very sensitive and, using two-liter sample volumes, has been shown capable of measuring TBT concentrations in surface waters down to approximately 1 ng/L. (Similar methods are used in other laboratories in the US and internationally to measure TBT in aqueous samples.) As a solvent extraction method the procedure is well suited to working with aqueous samples that might exert matrix interferences since problems with interfering constituents can be eliminated by the extraction of tributyltin out of the original sample matrix. The procedure is however labor intensive and a limited number of samples can be processed in a given period of time.

The benefit of the Fast TBT Method is that a greater throughput (i.e. production) can be achieved and results can be obtained in a shorter period of time after a sample is delivered to a laboratory. While there can be great benefits to this quick turnaround, if the results from the Fast TBT Method are not reproducible or consistent with the solvent extraction method, its value may prove to be limited. Its limited value would in part derive from the lack of acceptance of the procedure as an equivalent method by regulatory authorities. As a consequence, a comparison study between the Fast TBT Method and an accepted method(s) is needed to validate the capability of the method as a substitute for regulatory-recognized methods. Reported below is a summary of a limited inter-comparison round robin study between the Fast TBT method and the accepted Grignard derivitization procedure. The protocol, procedures, and results for the study are briefly outlined below. A more complete report is included as an appendix.

Experimental Design:

The comparison study was designed to assess whether the newly developed Fast TBT Method quantitatively measured TBT at levels consistent with the Grignard derivitization method. The null hypothesis to be tested was:

H_o: The Fast TBT Method measures TBT in aqueous samples at equivalent levels to the established method at a 95% confidence level (i.e. there is no difference between the Fast TBT Method and the established analytical procedure).

The hypothesis was tested for individual "groups" of samples and for all samples generated in the study. Statistical treatment is discussed in the appendix.

***Samples, Sample Handling, and Transport*:**

"Samples" for the study consisted of deionized, distilled water spiked to concentrations from < 10 ng/L TBT to approximately 10,000 ng/L TBT. Each sample used in the study was created in the environmental engineering laboratory at Old Dominion University in a single vessel/container and then split into identical aliquots into identical bottles and then distributed to the participating laboratories. Each laboratory analyzed the samples without filtering and without a preparatory digestion step. Each laboratory also made up and analyzed a (laboratory) blank.

Sample containers used throughout this study were 2-liter polycarbonate bottles. All sample bottles were washed with tap water, then methanol, and then 3 N HCl. After each step in the cleaning, bottles were copiously rinsed with high-purity, deionized, distilled water. All samples were transmitted to the participating laboratories in coolers containing ice packs or ice. Samples were sent by overnight mail or delivered directly by personnel at the laboratory generating the samples to the laboratories participating in the comparison study.

Summary of Results

Three laboratories participated in an Inter-calibration Exercise Round Robin Study to examine the capabilities of the Fast TBT Method. Two of the laboratories, the Virginia Institute of Marine Sciences (VIMS) and Old Dominion University (ODU) Oceanography Laboratory were in good agreement on standard reference samples analyzed in the critical regulatory discharge range (6.1 ng/L to 140 ng/L of TBT), see Figure R-1 below.

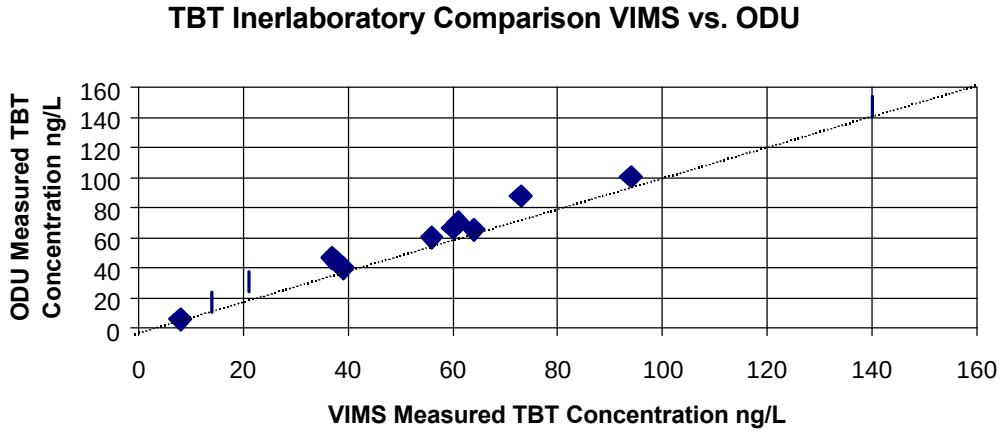

Figure R-1. Plot of VIMS to ODU Measured TBT Concentrations for Standard Reference Sample over the Test Range of 6.1 ng/L to 140 ng/L (the Critical Regulatory Discharge Range).

Good agreement was also observed in analysis of TBT for both VIMS and ODU from the 6.1 ng/L till about 4000 ng/L by both for the reference standards, see Figure R-2 below.

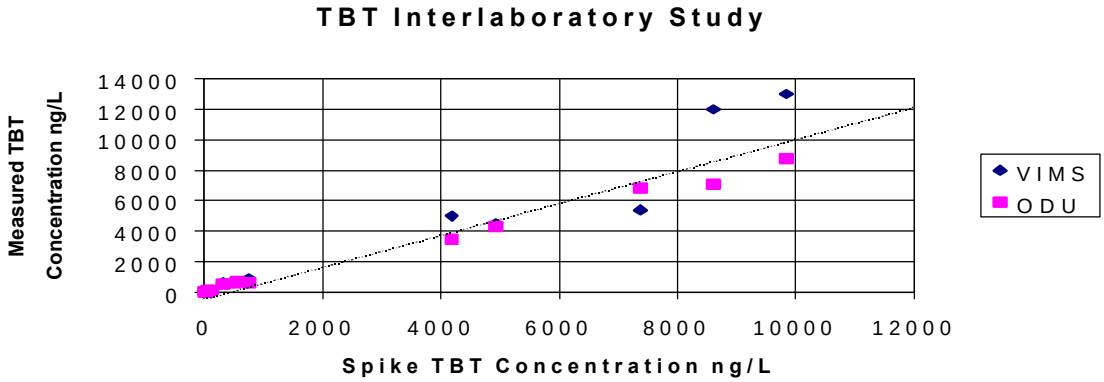

Figure R-2. Plot of VIMS to ODU Measured TBT Concentrations for Standard Reference Samples in ng/L for the Entire Range of Standard Reference Samples.

The results of this study confirm that for a simple matrix (deionized, distilled water) the methods produce comparable results (see appendix for additional discussion). Future research efforts will include evaluating the method in more complex matrices including samples generated at shipyards.

REFERENCES

Hodge, V.F, Seidel, S.L, Goldberg, E.D. 1979. Determination of tin (IV) and organotin compounds in natural waters. Anal. Chem., 51, 1256-1259.

Matthias, C.L., Bellama, J.M., Olson, G.L., Brinckman, F.E. 1986. Comprehensive method for determination of aquatic butyltin and butylmethyltin species at ultra trace levels using simultaneous hydrization/extraction with gas chromatography-flame photometric detection. Environ. Sci. Technol. 20, 609-615.

Sharron, D., Swami, K., Jansing, R.L. 1995. Rapid ultratrace analysis of tributyltin in aqueous matrixes by purge-and-trap gas chromatography with flame photometric detection. J. Assoc. Off. Anal. Chem., 78, 1317-1321.

Unger, M.A., MacIntyre, W.G., Greaves, J., Huggett, R.J. 1986. GC determination of butyltins in natural waters by flame photometric detection of hexyl derivatives with mass spectrometric confirmation. Chemosphere, 15, 461-470.

Valkirs, A.O., Seligman, P.F., Olson, G.J., Brinckman, F.E., Matthias, C.L., Bellama, J.M. 1987. Di- and tributyltin species in marine and estuarine waters. Inter-laboratory comparison of two ultratrace analytical methods employing hydride generation and atomic absorption or flame photometric detection. Analyst, 112, 17-21.

Appendix

Round Robin Interlaboratory Comparison of the Fast TBT Method with Established TBT Methods

Dr. Jaewan Yoon
Dr. Michael Champ

1.0 METHODS & EXPERIMENTAL DESIGN

A round robin comparison of TBT measurements by the Old Dominion University Laboratories, (ODU), the Virginia Institute of Marine Science (VIMS) Laboratory, and the Newport News Shipbuilding Laboratory (NNSBL) was conducted based on May 2000 sample data set to evaluate the sensitivity, precision, and accuracy of measurement methods using a standard statistical cross validation protocol with a 95% confidence level.

The main objective of the Round-Robin Intercalibration study is to determine the sensitivity, precision and accuracy of the three different contract laboratories. The Round Robin Intercalibration is designed in the following manner with a total of 20 spiked samples for analysis. The water samples were based on shipyard washdown wastewater treated to remove TBT, and then spiked with TBT standard certified reagent. To determine the sensitivity to various concentration ranges, reference sample(s) concentrations (RSC) were clustered into four subsets – VAWQS- (TBT NPDES Water Quality Standard for the State of Virginia), lower-, medium- and higher-ranges (Table 1.1).

VAWQS-range subset contained up to the second quartile (Q2: 0%-50%) of the entire reference sample(s) sample concentration range, and lower-range subset contained a remaining sample space between (70%-Q2), medium-range contained a sample space between (90%-70%) and higher-range contained a sample space between (100%-90%). Particularly, the medium-range subset was a cascade sample window of the higher-range subset, and designed to capture any localized, yet intrinsic gradient deflections that might occur between 800 ng/L and 8,000 ng/L TBT range to supplement the standard three-point validation characterization curve technique.

The Intercalibration is designed with samples clustered in four critical concentrations relative to the lower end of the VA NPDES Permit and CASRM Barge System Treatment Levels:

Table 1.1. Reference Sample(s) Concentration (RSC) Partitions in Subsets used in Intercalibration Exercises (May 2000).

Range Designation	Number of Samples	Spiked Sample TBT Concentration Range
VAWQS	10	6.1-57.8 ng/L
Lower	5	70-800 ng/L
Medium	3	4,000-7,000 ng/L
Higher	2	8,000–10,000 ng/L
	Total 20 Samples	6.1-10,000 ng/L

Total number of the sample size, n_{total}, to be used in this Round-Robin cross validation study is expressed by

$$n_{total} = 3\left[\sum_{VAWQS=1}^{i} n_{VAWQS} + \sum_{L=1}^{j} n_L + \sum_{M=1}^{k} n_M + \sum_{H=1}^{l} n_H\right]$$

$$(VAWQS = 1, 2, \ldots, i\ ;\ L = 1, 2, \ldots, j\ ;\ M = 1, 2, \ldots, k\ ;\ H = 1, 2, \ldots, l\ ;\ i \neq j \mid k \mid l\ ;\ j \neq k \mid l\ ;\ k = l)$$

where n_{VAWQS} = Number of the reference sample(s) concentration variations in the VAWQS-range subset (between 6.1 ng/L and 57.8 ng/L)
n_L = Number of the reference sample(s) concentration variations in the lower-range subset (between 70 ng/L and 800 ng/L)
n_M = Number of the reference sample(s) concentration variations in the medium-range subset (between 4,000 ng/L and 7,000 ng/L)
n_H = Number of the reference sample(s) concentration variations in the higher-range subset (between 8,000 ng/L and 10,000 ng/L)

Instead of partitioning up to the orthogonalized first quartile (Q1: 0%-25%), the VAWQS-range encompasses to the second quartile (Q2: 0%-50%) of the entire reference sample(s) sample concentration range to ensure the cross validation of the sensitivity of two TBT measurement methods. By using this approach, typical data hysteria and outliers prevalent at extreme tails (lower and upper) in a data distribution can be validly identified and buffered, and interpretation of cross-validation data distribution and subsets would not be skewed by possible localized extreme values. Also, this buffered VAWQS-range approach would facilitate a close examination of two TBT measurement methods whether methods can accurately detect low TBT concentrations since current Virginia DEQ's TBT water quality standard (WQS) for the State of Virginia is 50 ng/L.

2.0. RESULTS & DISCUSSION

Cross validation was conducted on several levels of accuracy, precision and bias of TBT measurement methods employed by three laboratories. A Student t-test based on the assumption of an unknown population variance (σ^2), both for equal and unequal σ^2 conditions, and the Central Limit Theorem (CLT) normal approximation were used to test this three way - cross validation experiment.

To ensure impartiality in statistical analysis and in the QAPP, the project quality assurance manager randomly reassigned laboratory identification aliases, i.e., Lab 1, Lab 2 and Lab3, to participating laboratories and the project quality assurance manager only has the access to the corresponding key lookup table for the aliases. Table 2.0 lists the original round-robin intercalibration TBT sample data sets used in this statistical analysis, and data sets were provided by the project quality assurance manager with laboratory identification aliases.

Table 2.0
Comparative TBT Concentrations Reported in CASRM Reference Standards for Intercalibration Exercises (May 2000).

Exercise Description	Reference Sample ID No.	TBT Spike TBT (ng/L)	LAB 1 TBT (ng/L)	Lab 2 TBT (ng/L)	Lab 3 TBT (ng/L)
Phase 1	1S	6.1	8	<7.0	13
	2S	17.2	14	17.4	13
	3S	22.1	21	30.8	14
	4S	32	39	40.1	18
	5S	39.3	37	46.9	7
	6S	59	61	70.4	9
Phase 2	11S	22.1	64	65.5	12
	12S	33.9	56	60.4	10
	13S	45.5	60	66.7	<1
	14S	57.8	73	87.8	14
	15S	89.7	94	100.7	31
	16S	120.4	140	147.7	46
Phase 3	21S	320	610	529.1	481
	22S	492	620	567.7	133
	23S	741	890	640.3	222
	24S	4180	5000	3484	938
	25S	4917	4500	4293	3289
	26S	7372	5400	6841	4922
	27S	8604	12,000	7086	4237
	28S	9832	13,000	8758	6961

2.1. Accuracy, Precision and Bias in Laboratory Measurements

For all four sample range subsets, VAWQS, lower-, medium-, and higher-range, measurements from three laboratories were compared for the level of deviation from the reference sample(s) concentration using the Youden plot (Youden, 1959) -- reference sample(s) sample concentration vs. laboratory measurement with two different methods as a variability component. (Figures 2.1.1-2.1.4) Statistical analysis was conducted in the Sun ES25000/StarFire High Performance Computing SMP Supercomputer cluster at the Old Dominion University using SAS 6.12 statistical analysis software on Sun Solaris 5.6 platform. Estimates from SAS Univariate procedure (SAS/PROC, 1990) are listed in Appendix I.SAS (SAS/STAT, 1990).

Accuracy is a function of both bias and precision. Bias measures systematic errors and precision measures the degree of scatter in the data. Methods that give accurate measurements have good precision and near zero bias. Inaccurate methods can have poor precision, unacceptable bias, or both. Once identified, bias can be removed by careful checks on experimental technique and equipment. Bias, hence, cannot be averaged out by making more measurements.

Table 2.1 summarizes comparison of accuracy, precision and bias of four TBT sample range subsets estimated from univariate descriptive statistics based on differences of each subset measurement from the reference sample(s) concentration. Individual summary for each sample range subset is also given in Sections 2.1.1 through 2.1.4.

2.1.1. VAWQS Range (6.1-59 ng/L) - Accuracy, Precision and Bias in TBT Intercalibration Exercise

Differential univariate statistics and corresponding accuracy, precision and bias statistics of VAWQS Range (6.1-59 ng/L) subset from Intercalibration Exercises (May 2000) are shown below in Table 2.1.1. All samples are normally distributed at a 95% level of confidence with Shapiro-Wilk W-statistics (Shapiro and Wilk, 1965) *p*-value greater than 0.05. Shapiro-Wilk W-statistic is the ratio of the best estimator of the variance (based on the square of a linear combination of the order statistics) to the usual corrected sum of squares estimator of variance. W ranges between zero and one, with small values of W leading to rejection of the null hypothesis of normality. Interpretation of accuracy, precision and bias is tallied below;

Table 2.1
Comparison of accuracy, precision and bias of four TBT sample range subsets

	Accuracy	Precision	Bias	Reproducibility	Outlying Run/RMSE
Lab 1	High	Good	Small	Good	Yes/Small
Lab 2	Good	Moderate	Small	Moderate	Yes/Moderate
Lab 3	Poor	Poor	Large	Poor	Yes/Large

Overall, accuracy and reproducibility of Lab 1 and Lab 2 measurements are within a 95% confidence interval for the lower TBT range (6.1-35 ng/L) but with a consistently higher than spike RSC measurement values beyond 35-59 ng/L range. It should be noted that those higher than spike RSCs were resulted from the Phase 2 Round-Robin exercise. Lab 3 exhibits poor accuracy and precision as well as large bias and low reproducibility of TBT RSC spikes used in the round-robin exercise in the VAWQS Range (6.1-59 ng/L). It is recommended that Lab 3 should review and calibrate the measurement method or procedure or both to correct the problem. Magnitudes of accuracy, precision and bias from three participating laboratories are also graphically represented in a Youden plot in Figure 2.1.1.

Table 2.1.1. Differential Univariate Statistics and Conformity Statistics of Three Laboratory TBT RSC Measurements VAWQS Range, 6.1-59 ng/L; N = 30; n = 10 per laboratory

	Lab 1	Lab 2	Lab 3
Mean	9.8	15.799	-22.4001
Variance (s^2)	199.8067	196.5576	380.5005
Standard Deviation	14.1353	14.0199	19.50642
Skewness	1.423406	0.842373	-0.11289
Kurtosis	1.989992	-0.07667	-1.39275
Coefficient of Variation	144.2377	88.73915	-87.0818
Standard Error of Mean	4.469974	4.433481	6.168472
Pr<W (Shapiro-Wilk)	0.059	0.2811	0.5552
Range	45.1	43.2	56.9
Bias (η-based)*	**8.7337**	**14.7327**	**-23.4664**
Precision	**14.1353**	**14.0199**	**19.50642**
RMSE	**16.60921**	**20.65212**	**29.05538**

Where * Berthouex *et al.*, (1994), Statistics for Environmental Engineers, CRC Press.

RMSE (Root Mean Square Error) = $\sqrt{\frac{\sum(\text{Measured}_i - \text{RSC}_i)}{n}}$, i = 1 to 10, and n=10

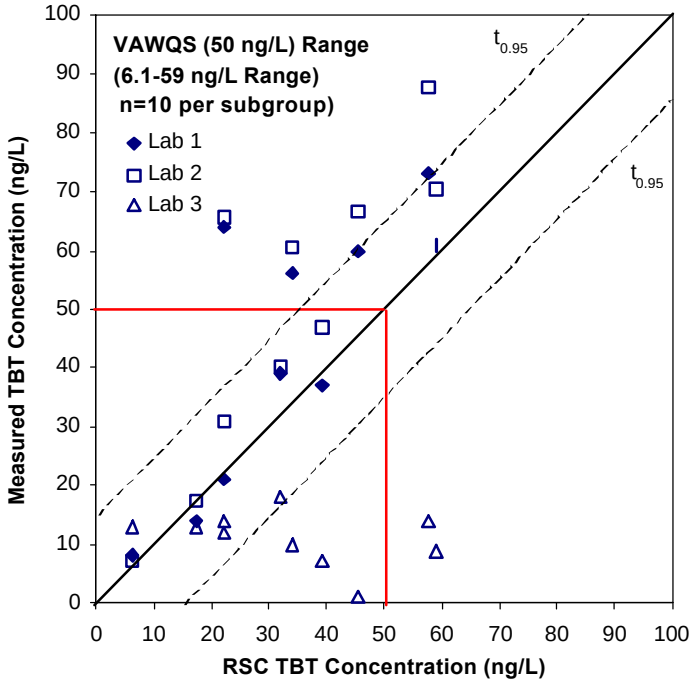

Figure 2.1.1. Accuracy, Precision and Bias Youden Plot of TBT Intercalibration Exercises (May 2000) with a 95% Confidence Interval for VAWQS Sample Range Subset, 6.1-59 ng/L; N = 30; n = 10 per laboratory

2.1.2. Lower Range (70-800 ng/L) - Accuracy, Precision and Bias in TBT Intercalibration Exercise

Differential univariate statistics and corresponding accuracy, precision and bias statistics of Lower Range (70-800 ng/L) subset from Intercalibration Exercises (May 2000) are shown below in Table 2.1.2. All samples are normally distributed at a 95% level of confidence with Shapiro-Wilk W-statistics *p*-value greater than 0.05.

Interpretation of accuracy, precision and bias is tallied below;

	Accuracy	Precision	Bias	Reproducibility	Outlying Run/RMSE
Lab 1	Good	Moderate	Moderate	Good	Absent/Small
Lab 2	Good	Moderate	Small	Good	Absent/Small
Lab 3	Moderate	Moderate	Large	Poor	Yes/Large

Accuracy and reproducibility of all Lab 1 and Lab 2 measurements are within a 95% confidence interval for the lower TBT range (70-800 ng/L). It should be noted the measurements between 300-750 ng/L that are higher than spike RSCs were resulted from the Phase 2 Round-Robin exercise. Lab 3 exhibits poor accuracy and precision as well as large bias and low reproducibility of TBT RSC spikes used in the Lower Range (70-800 ng/L). It is recommended that Lab 3 should review and calibrate the measurement method or procedure or both to correct the problem. Magnitudes of accuracy, precision and bias from three participating laboratories are also graphically represented in a Youden plot in Figure 2.1.2.

Table 2.1.2.
Differential Univariate Statistics and Conformity Statistics of Three Laboratory TBT RSC Measurements Lower Range, 70-800 ng/L; N = 15; n = 5 per laboratory

	Lab 1	**Lab 2**	**Lab 3**
Mean	118.18	44.48	-170.02
Variance (s^2)	13313.77	12641.93	72152.51
Standard Deviation	115.3853	112.4363	268.6122
Skewness	0.723597	0.407165	-0.23058
Kurtosis	0.004589	1.23657	-1.19264
Coefficient of Variation	97.63523	252.7796	-157.989
Standard Error of Mean	51.60188	50.28306	120.127
Pr<W (Shapiro-Wilk)	0.5189	0.8691	0.7624
Range	285.7	309.8	680
Bias (η-based)	**120.6333**	**46.93333**	**-167.567**
Precision	**115.3853**	**112.4363**	**268.6122**
RMSE	**110.9449**	**77.75608**	**208.1211**

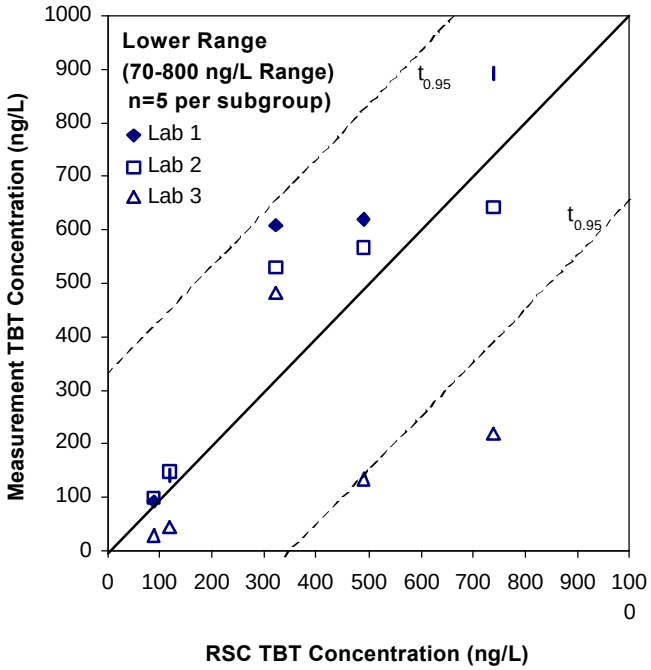

Figure 2.1.2. Accuracy, Precision and Bias Youden Plot of TBT Intercalibration Exercises (May 2000) with a 95% Confidence Interval for Lower Sample Range Subset, 70-800 ng/L; N = 15; n = 5 per laboratory

2.1.3. Medium Range (4000-7000 ng/L) - Accuracy, Precision and Bias in TBT Intercalibration Exercise

Differential univariate statistics and corresponding accuracy, precision and bias statistics of Medium Range (4000-7000 ng/L) subset from Intercalibration Exercises (May 2000) are shown below in Table 2.1.3. All samples are normally distributed at a 95% level of confidence with Shapiro-Wilk W-statistics *p*-value greater than 0.05. Interpretation of accuracy, precision and bias is tallied below;

	Accuracy	Precision	Bias	Reproducibility	Outlying Run/RMSE
Lab 1	Moderate	Moderate	Small	Good	Absent/Moderate
Lab 2	Good	Good	Small	Good	Absent/Small
Lab 3	Moderate	Poor	Moderate	Moderate	Yes/Large

Accuracy and reproducibility of all laboratories are within a 95% confidence interval for the Medium TBT range (4000-7000 ng/L) except one outlying run from Lab 3. Particularly, Lab 2 exhibits best accuracy and precision as well as smallest bias and high reproducibility of TBT

RSC spikes used in the Medium Range (4000-7000 ng/L). For Lab 3, it is further recommended to identify the cause of outlying run at 4180 ng/L and proximity ranges to correct the problem. Magnitudes of accuracy, precision and bias from three participating laboratories are also graphically represented in a Youden plot in Figure 2.1.3.

Table 2.1.3
Differential Univariate Statistics and Conformity Statistics of Three Laboratory TBT RSC Measurements
Medium Range, 4000-7000 ng/L; N = 9; n = 3 per laboratory

	Lab 1	**Lab 2**	**Lab 3**
Mean	-523	-617	-2440
Variance (s^2)	1957243	6843	651324
Standard Deviation	1399.015	82.72243	807.0465
Skewness	-0.339	0.378065	0.05575
Coefficient of Variation	-267.498	-13.4072	-33.0757
Standard Error of Mean	807.7217	47.75982	465.9485
Pr<W (Shapiro-Wilk)	0.8743	0.8597	0.9779
Range	2792	165	1614
Bias (η-based)	**-165**	**-259**	**-2082**
Precision	**1399.015**	**82.72243**	**807.0465**
RMSE	**688.1187**	**339.9637**	**1384.321**

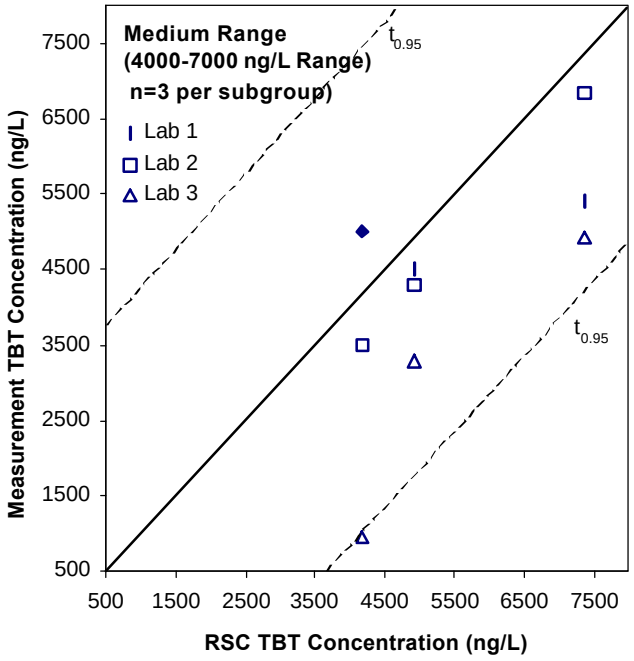

Figure 2.1.3. Accuracy, Precision and Bias Youden Plot of TBT Intercalibration Exercises (May 2000) with a 95% Confidence Interval for Medium Sample Range Subset, 4000-7000 ng/L; N = 15; n = 5 per laboratory

2.1.4. Higher Range (8000-10000 ng/L) - Accuracy, Precision and Bias in TBT Intercalibration Exercise

Differential univariate statistics and corresponding accuracy, precision and bias statistics of Higher Range (8000-10000 ng/L) subset from Intercalibration Exercises (May 2000) are shown below in Table 2.1.4. All samples are normally distributed at a 95% level of confidence with Shapiro-Wilk W-statistics *p*-value greater than 0.05. Interpretation of accuracy, precision and bias is tallied below;

	Accuracy	Precision	Bias	Reproducibility	Outlying Run/RMSE
Lab 1	Moderate	Moderate	Moderate	Moderate	Absent/Moderate
Lab 2	Good	Good	Small	Good	Absent/Small
Lab 3	Moderate	Poor	Moderate	Moderate	Yes/Large

Accuracy and reproducibility of all laboratories are within a 95% confidence interval for the Higher TBT range (8000-10000 ng/L) except one outlying run from Lab 3. Particularly, Lab 2 exhibits best accuracy and precision as well as smallest bias and high reproducibility of TBT RSC spikes used in the Higher Range (8000-10000 ng/L). For Lab 3, it is further recommended to identify the cause of outlying run at 8640 ng/L and proximity ranges to correct the problem. Magnitudes of accuracy, precision and bias from three participating laboratories are also graphically represented in a Youden plot in Figure 2.1.4.

Table 2.1.4. Differential Univariate Statistics and Conformity Statistics of Three Laboratory TBT RSC Measurements Higher Range, 8000-10000 ng/L; N = 6; n = 2 per laboratory

	Lab 1	**Lab 2**	**Lab 3**
Mean	3282	-1296	-3619
Variance (s^2)	25992	98568	1119008
Standard Deviation	161.2203	313.9554	1057.832
Coefficient of Variation	4.912259	-24.225	-29.2299
Standard Error of Mean	114	222	748
Pr<W (Shapiro-Wilk)	> 0.05	> 0.05	> 0.05
Range	228	444	1496
Bias (η-based)	**3390.867**	**-1187.13**	**-3510.13**
Precision	**161.2203**	**313.9554**	**1057.832**
RMSE	**1468.64**	**588.0306**	**1652.674**

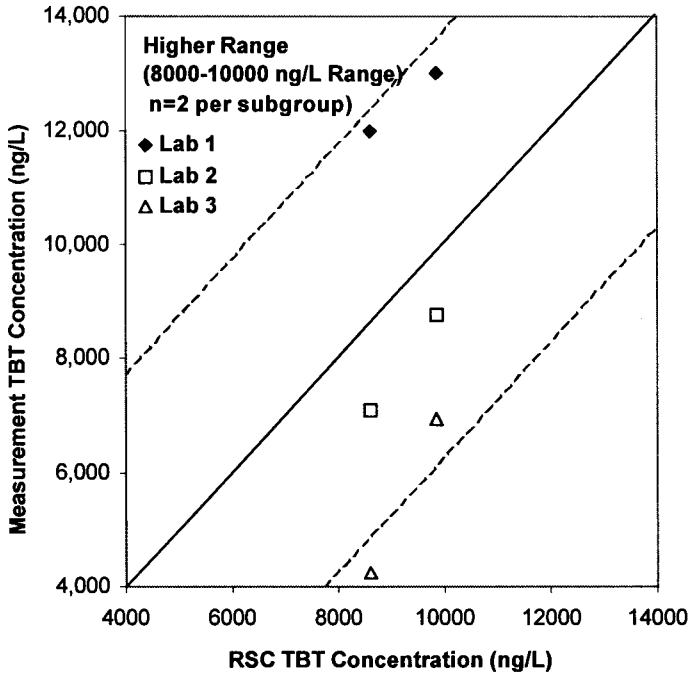

Figure 2.1.4. Accuracy, Precision and Bias Youden Plot of TBT Intercalibration Exercises (May 2000) with a 95% Confidence Interval for Higher Sample Range Subset, 4000-7000 ng/L; N = 6; n = 2 per laboratory

2.2. Tests of Hypothesis on Means Among Laboratory Measurements

Cross validation was conducted on several levels of test of hypothesis (T.H.) on means of measured TBT concentrations, both singular and multiple per each subset of the reference sample(s) sample concentrations for each range subset, by three laboratories. A Student t-test with a 95% confidence level.($\alpha = 0.05$) based on the assumption of an unknown population variance (σ^2), both for equal and unequal σ^2 conditions, and the Central Limit Theorem (CLT) normal approximation were used to test this three way - cross validation experiment.

To ensure impartiality in statistical analysis and in the QAPP, the project quality assurance manager randomly reassigned laboratory identification aliases, i.e., Lab 1, Lab 2 and Lab3, to participating laboratories and the project quality assurance manager only has the access to the corresponding key lookup table for the aliases. Table 2.0 lists the original round-robin intercalibration TBT sample data sets used in this statistical analysis, and data sets were provided by the project quality assurance manager with laboratory identification aliases.

2.2.1. Studentized t-statistics Comparison of Laboratory Means and Variances to RSC in Subgroups.

Student t-statistics with a 95% confidence were estimated for equal ($\sigma_{RSC}^{2} = \sigma_{Lab}^{2}$) and unequal ($\sigma_{RSC}^{2} \neq \sigma_{Lab}^{2}$) population variance assumption so that each assumption's sensitivity can be throughly tested. For the unequal population variance assumption, to further analyze the influence of the degree of freedom in Student t-statistics, more conservative Satterthwaite, Cochran and Cox approximation method (Cochran *et al*., 1950, Lee and Gurland, 1975) was used in addition to the standard unequal population variance t-statistics approximation. Test of hypothesis on sample variance was also conducted using F-statistics.

Singular hypothesis, i.e., Spike vs. each subset was set as two- and one-way comparison;

H_0: μ_{TBT} [Subset/RSC] = μ_{TBT} [Subset/Laboratory (Lab 1|Lab 2|Lab 3)]
H_a: μ_{TBT} [Subset/RSC] ≠ μ_{TBT} [Subset/Laboratory (Lab 1|Lab 2|Lab 3)] (Dual T.H.)
μ_{TBT} [Subset/RSC] > μ_{TBT} [Subset/Laboratory (Lab 1|Lab 2|Lab 3)] (Singular T.H.)
μ_{TBT} [Subset/RSC] < μ_{TBT} [Subset/Laboratory (Lab 1|Lab 2|Lab 3)] (Singular T.H.)

Hypothesis for F-statistics was set as two- and one-way comparison at a 95% confidence level;

H_0: σ_{TBT}^{2} [Subset/RSC] = σ_{TBT}^{2} [Subset/Laboratory (Lab 1|Lab 2|Lab 3)]
H_a: σ_{TBT}^{2} [Subset/RSC] ≠ σ_{TBT}^{2} [Subset/Laboratory (Lab 1|Lab 2|Lab 3)] (Dual T.H.)
σ_{TBT}^{2} [Subset/RSC] > σ_{TBT}^{2} [Subset/Laboratory (Lab 1|Lab 2|Lab 3)] (Singular T.H.)
σ_{TBT}^{2} [Subset/RSC] < σ_{TBT}^{2} [Subset/Laboratory (Lab 1|Lab 2|Lab 3)] (Singular T.H.)

Student t-statistic with an equal population variance assumption ($\sigma_{BRC}^{2} = \sigma_{Lab}^{2}$)

$$T_0 = \frac{\bar{x}_{BRC} - \bar{x}_{Lab}}{s_p \sqrt{\frac{1}{n_{BRC}} + \frac{1}{n_{Lab}}}} \quad \text{and} \quad s_p = \sqrt{\frac{(n_{BRC}-1)s_{BRC}^2 + (n_{Lab}-1)s_{Lab}^2}{n_{BRC}+n_{Lab}-2}}$$

where s_p is a spooled sample standard deviation

Student t-statistic with an unequal population variance assumption ($\sigma_{BRC}^{2} \neq \sigma_{Lab}^{2}$)

$$T_0 = \frac{\bar{x}_{BRC} - \bar{x}_{Lab}}{\sqrt{\frac{s_{BRC}^2}{n_{BRC}} + \frac{s_{Lab}^2}{n_{Lab}}}}$$ with an integer(ν) degree of freedom

$$\text{where } n = \left[\frac{\left(\frac{s_{BRC}^2}{n_{BRC}} + \frac{s_{Lab}^2}{n_{Lab}} \right)^2}{\frac{\left(\frac{s_{BRC}^2}{n_{BRC}} \right)^2}{n_{BRC}+1} + \frac{\left(\frac{s_{Lab}^2}{n_{Lab}} \right)^2}{n_{Lab}+1}} \right] - 2$$

Summary of test of hypothesis is shown in Table 2.2.1 with associated *p*-values (= Prob > T|F'). Estimates from SAS (SAS/STAT, 1990) T-test procedure are listed in Appendix II.

Table 2.2.1. Summary of Singular and Dual Student t-statistics for Round-Robin Inter-Laboratory Calibration study ($\alpha = 0.05$)

VAWQS Range (6.1-57.8 ng/L)	t-statics	Prob>|T| (Equal σ^2)	Prob>|T| (Unequal σ^2)	Prob>|T| (Cochran)	Prob>F'
Spike vs. Lab 1	-1.0792	0.2947	0.2958	0.3086	0.4164
Spike vs. Lab 2	-1.6206	0.1225	0.1248	0.1396	0.2636
Spike vs. Lab 3	3.9502	0.0009	0.0026	0.0034	0.0006
Lower Range (70-800 ng/L)					
Spike vs. Lab 1	-0.6050	0.5619	0.5629	0.5778	0.6632
Spike vs. Lab 2	-0.2682	0.7953	0.7954	0.8018	0.8948
Spike vs. Lab 3	1.1608	0.2792	0.2837	0.3103	0.4675
Medium Range (4000-7000 ng/L)					
Spike vs. Lab 1	0.5233	0.6284	0.6483	0.6530	0.1357
Spike vs. Lab 2	0.4414	0.6818	0.6818	0.7021	0.9529
Spike vs. Lab 3	1.1608	0.2792	0.2837	0.3103	0.4675
Higher Range (8000-10000ng/L)					
Spike vs. Lab 1	-4.1448	0.0536	0.0612	0.1507	0.8702
Spike vs. Lab 2	1.2495	0.3379	0.3530	0.4297	0.8066
Spike vs. Lab 3	1.6202	0.1805	0.1834	0.2466	0.8210

In summary, with 95% confidence,

i) there is no sufficient evidence indicating that RSC concentration mean is significantly different from all four subgroup means except the mean and variance of Lab 3 measurements in the VAWQS range (6.1-57.8 ng/L), and

ii) Furthermore, from the result from the one-way T.H. on the mean of Lab 3 measurements in the VAWQS range indicates that there is sufficient evidence that the mean of Lab 3 measurements is significantly smaller than the mean of RSC in the VAWQS range (*p*-value = 0.0018|0.0052|0.0072) with 95% confidence level.

2.2.2. Multiple Mean Comparison (MMC) of Laboratory Means to RSC in Subgroups.

Pairwise and order of the magnitude mean concentration comparisons with a 95% confidence were conducted to find the similarity among three laboratory TBT measurements. Standard Fisher's least significant difference (LSD) (Fisher, 1958) and Student-Newman-Keuls (SNK) (Keuls, 1952; Newman, 1939) with a 95% confidence level ($\alpha = 0.05$) were used for this comparison for all four subgroups. Fisher's LSD test is most widely accepted for pairwise mean comparisons, and the α-level of Fisher's LSD test is valid for independent event, i.e., orthogonal, comparisons. Student-Newman-Keuls (SNK) is based on a modification of the Tukey's Studentized Range test, and uses different critical values for comparing pairwise means in comparison to the Fisher's LSD test. SNK test is regarded generally more conservative than the Fisher's LSD test, and SNK test will be used to validate pairwise mean comparisons from initial Fisher's LSD test.

Fisher's LSD: $$LSD = t_{\frac{\alpha}{2}} \sqrt{MS_E{}^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

where MS_E = Mean squared sum of error within samples
n_i = Number of samples used in the pairwise comparison
n_j = Number of samples used in the pairwise comparison

SNK's Critical value: $$W_r = q_{\alpha}(r, \nu) \sqrt{MS_E{}^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

where $q_\alpha(r, \nu)$ = Critical value of the Studentized range
MS_E = Mean squared sum of error within samples
ν = Degree of freedom for MS_E
n_i = Number of samples used in the pairwise comparison
n_j = Number of samples used in the pairwise comparison

Summary of Multiple Mean Comparison (MMC) is shown in Table 2.2.2. Estimates from SAS (SAS/STAT, 1990) ANOVA procedure are listed in Appendix III.

Table 2.2.2. Summary of Multiple Mean Comparison (MMC) of Laboratory Means to RSC in Subgroups for Round-Robin Inter-Laboratory Calibration study ($\alpha = 0.05$)

VAWQS Range (6.1-57.8 ng/L)	Spike	Lab 1	Lab 2	Lab 3
Spike	*	Same	Same	Different
Lab 1	Same	*	Same	Different
Lab 2	Same	Same	*	Different
Lab 3	Different	Different	Different	*
Lower Range (70-800 ng/L)				
Spike	*	Same	Same	Same
Lab 1	Same	*	Same	Same
Lab 2	Same	Same	*	Same
Lab 3	Same	Same	Same	*
Medium Range (4000-7000 ng/L)				
Spike	*	Same	Same	Same
Lab 1	Same	*	Same	Same
Lab 2	Same	Same	*	Same
Lab 3	Same	Same	Same	*
Higher Range (8000-10000ng/L)				
Spike	*	Same	Same	Same
Lab 1	Same	*	Same	Different
Lab 2	Same	Same	*	Same
Lab 3	Same	Different	Same	*

* Intrinsic Null Comparison

In summary, there is sufficient evidence indicating that;

i) The mean of TBT measurements made by Lab 3 is significantly differ from the means of Spike, Lab 1, and Lab 2 ($\mu_{Lab3} < \mu_{Spike|Lab1|Lab2}$) in VAWQS (6.1-57.8 ng/L) range subgroup and

ii) The mean of TBT measurements made by Lab 1 is significantly differ from the mean of Lab 3 ($\mu_{Lab1} > \mu_{Lab3}$) in Lower range (70-800 ng/L) subgroups with 95%

confidence.

iii) For Lower and Medium range subgroups, all means from three laboratories are the same to the means of RSC spikes.

2.3. CONCLUSIONS AND RECOMMENDATIONS

i) Based on Univariate and Accuracy, Precision, and Bias analyses, measurements from Lab 3 exhibited a significantly lower RSC reproducibility compared to Lab 1 and Lab 2 in VAWQS (6.1-57.8 ng/L) range subgroup at 95% confidence level. It is recommended that Lab 3 should review and calibrate the measurement method and/or laboratory procedure to correct the precision and bias problems.

ii) Measurements from Lab 1 and 2 are consistently within 95% confidence level for all four subgroups – VAWQS, Lower-, Medium- and Higher-ranges. Lab 1 was able to reproduce RSCs most closely among three laboratories in VAWQS range subgroup. On the other hand, Lab 2 reproduced RSCs most closely in a consistent manner for Lower-, Medium- and Higher-range subgroups.

iii) Measurement means of Lab 3 are statistically equal to RSC means in Lower-, Medium- and Higher-range subgroups at 95% confidence level even though measurements are consistently lower than RSCs. This is due to the magnitude of RSC concentration ranges in Lower-, Medium- and Higher-range subgroups.

iv) A pairwise multiple mean comparison (MMC) results indicated that the mean and variance of measurements from Lab 3 significantly differs from those of RSC, Lab 1 and Lab2 in VAWQS range subgroup at 95% confidence level, which supports the conclusion from Univariate and Accuracy, Precision, and Bias analysis resulted in i).

v) Means and variances of measurements from all three laboratories and RSCs are equal for Lower-, Medium-range subgroups at 95% confidence level.

vi) Means of measurements from Lab 1 and lab 3 are different from each other, with the mean of Lab 1 is greater than the mean of Lab 3 at 95% confidence level in Higher range subgroup. However, both means of Lab 1 and Lab 3 are equal to the mean of RSC in Higher range subgroup.

vii) It should be noted that TBT measurements from Lab 1 and Lab 2 (excluding Lab 3) were persistently higher yet consistently than RSCs for samples taken in Phase 2 Round-Robin Inter-laboratory exercise that were encompassing the VAWQS range subgroup (6.1-57.8 ng/L). Specific RSCs corresponding to elevated laboratory measurements are 22.1, 33.9, 45.5, and 57.8 ng/L. It is suspected that

all samples used in Phase 2 Round-Robin Inter-laboratory exercise were possibly contaminated in the source level. It is recommended to repeat measurements for the corresponding RSCs again for further verification.

In conclusion, laboratory method and/or procedure used in Lab 1 and Lab 2 are acceptable at 95% confidence level for future TBT measurements. In contrast, laboratory method and/or procedure used in Lab 3 is not acceptable at 95% confidence level, and should be corrected before engaging future TBT analysis.

References

Cochran, W.G. and Cox, G.M., (1950), Experimental Designs, Jon Wiely and Sons.

Kuels, M., (1952), "The Use of the Studentized Range in Connection with an Analysis of Variance," *Euphytica*, 1:112-122.

Lee, A.F.S. and Gurland, J., (1975), "Size and Power of Tests for Equality of Means of Two Normal Populations with Unequal Variances," *J. American Statistical Association*, 70:933-941.

Fisher, R.A., (1958) Statistical Methods for Research Workers, 13th

Newman, D., (1939), "The Distribution of the Range in Samples from a Normal Population, Expressed in Terms of an Independent Estimate of Standard Deviation," *Biometrika*, 31:20-30.

Shapiro, S.S. and Wilk, M.B. (1965), "An Analysis of Variance Test for Normality (complete samples)," *Biometrika*, 5:591-611.

SAS/PROC, (1990), Statistical Analysis Systems/ Procedure Guide, Version 6, Third Edition, SAS Institute.

SAS/STAT, (1990), Statistical Analysis Systems/STAT User's Guide, Version 6, Fourth Edition, Vol. 2, SAS Institute.

Youden, W.J., (1959), "Graphical Diagnosis of Interlaboratory Test Results," *Industrial Quality Control*, 1959:24-28.