An Intercomparison of Aircraft Instrumentation for Tropospheric Measurements of Carbonyl Sulfide, Hydrogen Sulfide, and Carbon Disulfide

GERALD L. GREGORY, 1 DOUGLAS D. DAVIS, 2 DONALD C. THORNTON, 3 JAMES E. JOHNSON, 4 ALAN R. BANDY, 3 ERIC S. SALTZMAN, 5 MEINRAT O. ANDREAE, 6 AND JOHN D. BARRICK 1

This paper reports results of NASA's Chemical Instrumentation and Test Evaluation (CITE 3) during which airborne measurements for carbonyl sulfide (COS), hydrogen sulfide (H2S), and carbon disulfide (CS2) were intercompared. Instrumentation included a gas chromatograph using flame photometric detection (COS, H2S, and CS2), a gas chromatograph using mass spectrometric detection (COS and CS2), and the Natusch technique (H2S). The measurements were made over the Atlantic Ocean east of North and South America during flights from NASA's Wallops Flight Center, Virginia, and Natal, Brazil, in August/September 1989. Most of the intercomparisons for H2S and CS2 were at mixing ratios <25 pptv and <10 pptv, respectively, with a maximum mixing ratio of about 100 pptv and 50 pptv, respectively. Carbonyl sulfide intercomparisons were at mixing ratios between 400 and 600 pptv. Measurements were intercompared from data bases constructed from time periods of simultaneous or overlapping measurements. Agreement among the COS techniques averaged about 5%, and individual measurements were generally within 10%. For H2S and at mixing ratio >25 pptv, the instruments agreed on average to about 15%. At mixing ratios <25 pptv the agreement was about 5 pptv. For CS2 (mixing ratios <50 pptv), two techniques agreed on average to about 4 pptv, and the third exhibited a bias (relative to the other two) that varied in the range of 3-7 pptv. CS2 mixing ratios over the ocean east of Natal as measured by the gas chromatograph-mass spectrometer technique were only a few pptv and were below the detection limits of the other two techniques. The CITE 3 data are used to estimate the current uncertainty associated with aircraft measurements of COS, H2S, and CS2 in the remote troposphere.

INTRODUCTION

Sulfur gases and their reaction products play important roles in the chemistry of the global troposphere and in the biogeochemical sulfur cycle [e.g., Cullis and Hirschler, 1980; Rodhe and Isakson, 1980; Freneny et al., 1983; Toon et al., 1987; Saltzman and Cooper, 1988; Bates et al., 1987, 1990; Andere, 1990; Andere and Jaeschke, 1993]. Major sulfur gases include sulfur dioxide (SO2), dimethyl sulfide (DMS), hydrogen sulfide (H2S), carbon disulfide (CS2), and carbonyl sulfide (COS). Atmospheric photochemistry results in the oxidation of reduced sulfur gases (DMS, H2S, CS2, and COS) to SO2, methanesulfonic acid (MSA) and sulfuric acid. MSA and sulfuric acid are incorporated into the atmospheric aerosol and represent the dominant source of aerosol particles in the unpolluted marine atmosphere [National Aeronautics and Space Administration (NASA), 1984; Andere, 1990]. Important scientific issues in atmospheric sulfur chemistry include (1) the relative role of the various sources and sinks (natural and anthropogenic) of the major sulfur gases, (2) the rates and pathways of atmospheric oxidation of the sulfur species to SO2 and the production of aerosols, (3) the vertical distribution of sulfur gases and aerosols and their transport between the mixed layer and free troposphere, (4) the role of sulfur in acid rain, and (5) the impact of sulfate aerosol production (stratosphere and troposphere) as it relates to the Earth's energy balance and climate [Charlson et al., 1987].

The principal anthropogenic sulfur compound is sulfur dioxide, which is predominantly emitted during fossil fuel burning (about 70-100 Tg S/yr [Cullis and Hirschler, 1980]). Volcanoes are the major natural direct source of SO2, with an emission rate on the order of 10 Tg S/yr [Berresheim and Jaeschke, 1983]. The largest natural sulfur source to the atmosphere is the emission of dimethyl sulfide from the oceans, which amounts to some 15-40 Tg S/yr [Bates et al., 1987; Andere, 1990; Erickson et al., 1990]. Terrestrial biogenic sulfur emissions are predominantly in the form of hydrogen sulfide and dimethyl sulfide; the total flux of reduced sulfur gases from terrestrial ecosystems is about 4 Tg S/yr [Andere and Jaeschke, 1993]. The oceans are also a source of COS, H2S, and CS2, but with emission rates at least a factor of 10 lower than observed for DMS [Toon et al., 1987]. The natural flux of CS2 into the atmosphere is relatively small, and anthropogenic sources dominate CS2 production near industrial regions. DMS, H2S, and CS2 atmospheric oxidation are important sources for SO2. As a result of oxidation, H2S and CS2 have only a limited lifetime (hours) in the marine mixed layer [Andere, 1985].

Carbonyl sulfide is the longest lived (lifetime of several years) and thus the most abundant sulfur trace gas in the free troposphere.
troposphere (500 pptv compared to less than 100 pptv for SO$_2$ and less than 10 pptv for CS$_2$, H$_2$S, or DMS) [Torres et al., 1980; Carroll, 1985; Johnson and Harrison, 1986; Bingemer et al., 1990]. Carbonyl sulfide is emitted by the oceans [Rasmussen et al., 1982; Ferek and Andreae, 1983, 1984; Turner and Liss, 1985; Brasseur et al., 1990] and is also an important secondary product of atmospheric CS$_2$ oxidation [Kurkko, 1978; Sze and Ko, 1979] as well as being emitted during biomass burning and fossil fuel combustion. An estimated total oceanic COS emission of 0.1–0.5 Tg S/yr represents probably the largest single source of COS and may account for about one-third of the total source of tropospheric COS [Khalil and Rasmussen, 1984]. As a result of the long lifetime of COS, it is important to lower stratospheric sulfate aerosol production via oxidation mechanisms [Crutzen, 1976; Toon et al., 1979; Turco et al., 1980].

The various emission rates of DMS, CS$_2$, and H$_2$S and their subsequent oxidation to SO$_2$ and COS are not well documented. A major uncertainty in sulfur budget studies has been the validity of the various sulfur gas measurements. With emphasis on the marine environment, research has focused on the emission rate of the various sulfur gases as a function of seawater composition and meteorological conditions; the conversion of the various sulfur gases to SO$_2$, COS, and other sulfur compounds (gases and aerosols); and the mechanisms by which reaction products are transported throughout the global troposphere. As a result, several techniques have been used for the measurement of these sulfur gases. The question arises as to the validity of sulfur gas measurements at tropospheric concentrations which are in the parts-per-trillion (pptv) range. In the case of COS with a global background mixing ratio of about 500 pptv, a high level of measurement precision is required to account for predicted or measured COS hemispherical differences and to conduct meaningful sulfur budget studies.

As part of the NASA Tropospheric Chemistry Program, a series of field intercomparisons have been initiated to evaluate the state-of-the-art capability for measuring key tropospheric species [McNeal et al., 1983; Hoell et al., 1984; Gregory et al., 1985; Beck et al., 1987]. These intercomparisons, designated as Chemical Instrumentation Test and Evaluation (CITE), are conducted as part of NASA’s Global Tropospheric Experiment (GTE). The primary objective of the first intercomparison, GTE/CITE 1, was the evaluation of the capability for measurements of background levels of carbon monoxide (CO), nitric oxide (NO), and the hydroxyl radical (OH) [Hoell et al., 1984, 1985a, b]. CITE 2 extended the intercomparisons to the other major nitrogen gases, namely, nitrogen dioxide (NO$_2$), nitric acid (HNO$_3$), and peroxyacetyl nitrate (PAN) [Gregory et al., 1990a, b, c, d]. The objectives of CITE 3 were (1) to evaluate current instrumentation for their ability to make reliable aircraft measurements of the major sulfur gases and (2) to determine, in a predominantly marine environment, the abundance and distribution of the major sulfur gases over a wide range of atmospheric conditions.

This paper reports the results from CITE 3, during which airborne measurements for COS, H$_2$S, and CS$_2$ using instruments of different detection principles were intercompared. The measurements were made from the Wallops Electra aircraft during flights over the Atlantic Ocean during August and September 1989 from Wallops Island, Virginia, and Natal, Brazil. Intercomparison results for SO$_2$ and DMS are the subject of companion papers. The abundance and distribution of sulfur species in the marine atmosphere is the subject of numerous papers included in this issue of the Journal of Geophysical Research.

**EXPERIMENTAL DESCRIPTION**

**Instrumentation**

Table 1 summarizes the instruments which were intercompared. For COS and CS$_2$, three fundamentally different detection principles are represented: (1) gas chromatograph–flame photometric detection (GC), (2) gas chromatograph–mass spectrometric detection (MS), and (3) gas chromatograph/fluorination–electron capture detection (FLUOR). All three techniques use gas chromatography for sulfur species separation prior to detection. These instruments did not measure both COS and CS$_2$ during all flights, since during some flights, the instruments participated in the SO$_2$, H$_2$S, and/or DMS intercomparisons. (See Table 1 for a list of the sulfur gases measured by the instruments.) For example, the gas chromatograph–mass spectrometric technique measured COS only during the ferry flights between Wallops Island and Natal. H$_2$S intercomparison measurements were made with three instruments: gas chromatograph–flame photometric and two applications of the Natusch technique.

The notations in the last two columns of Table 1 are used in the paper to identify the instrumentation. A brief description of each instrument and its operation is given below. Detailed descriptions of the instruments are found in the references and companion papers. The configuration of the instrumentation aboard the aircraft is discussed in the overview paper [Hoell et al., this issue].

**Gas chromatograph–flame photometric (COS, H$_2$S, and CS$_2$)**. Sulfur gases in the incoming air stream are preconcentrated in a Teflon trap cooled with liquid argon. After preconcentration for several minutes (typically, 3 min during CITE 3), the trap contents are volatilized (via heating) into a carrier gas (air). The sulfur gases in the carrier gas are separated by gas chromatography and then analyzed flame photometrically. To optimize the detection of the various sulfur gases and to minimize analysis time and the effects of interferences, separate columns are used for the various sulfur gases. For CITE 3, COS and H$_2$S measurements were made with one column, DMS and CS$_2$ with a second column, and SO$_2$ with a third column. During CITE 3, a two-column measurement scenario was employed resulting in about a 10- to 12-min sample frequency for each of two gases. Inflight calibration was performed at frequent intervals (one to two per hour) using standard addition of the sulfur gas of interest near the sample inlet. The precision of the CITE 3 measurements was of the order of 40, 25, 12, and 10% for mixing ratios of 20, 50, 100, and 500 pptv, respectively. (The precision stated in Table 1 is that for a 50-pptv mixing ratio measurement.) Accuracy (primary standard) is of the order of 20%. The reader is referred to the references for further details [Bandy et al., 1981; Goldberg et al., 1981; Maroulis and Bandy, 1980; Maroulis et al., 1977; Torres et al., 1980].

**Gas chromatograph–mass spectrometer (COS and CS$_2$)**. Sulfur gases in the incoming air stream are also preconcentrated in a Teflon trap cooled with liquid argon for several minutes (typically, 3 min during CITE 3). The trap contents are then volatilized (via heating) into a carrier gas.
Separation of the sulfur gases in the carrier gas occurs by gas chromatography, and the separated sulfur gases are then analyzed with a quadrupole mass spectrometer operating in a single ion mode. An isotopically labeled variant of the gas being measured is constantly added (near the sample inlet) to the incoming atmospheric sample. Since the mass spectrometer can separately and simultaneously monitor the labeled (standard) and unlabeled (sample) species, a standard addition calibration is included with each measurement. Thus, sample losses that may occur in the inlet of the instrument are accounted for, as is any variation in the sensitivity of the mass spectrometer. Typically, about 3 to 4 min (start of trap heating to completion of analysis) are required for each sulfur gas measurement. Since the same instrument was used to measure multiple sulfur gases (generally two to three), the measurement frequency for any individual sulfur gas was about one sample every 10 to 12 min. The precision of the measurement is 10, 5, 3, and 1% for COS and CS$_2$ mixing ratios of 20, 50, 100, and 500 pptv, respectively. Accuracy (primary standard) is of the order of 20%. Further details are given in the references [Lewin et al., 1987; Bandy et al., 1985].

Gas chromatograph/fluorination-electron capture (COS and CS$_2$). Sulfur compounds in the incoming air sample are separated by gas chromatography and then fluorinated with F$_2$ (200 ppmv) using a heated Ag catalyst. The fluorination product, presumably SF$_6$, is then measured using an electron capture detector. The F$_2$ stream is generated from a permeation source, and excess F$_2$ is removed by conversion to HF by reaction with H$_2$ on a heated Pd catalyst. The Pd catalyst also destroys any response from halocarbons, making the system sulfur specific. Cryogenic preconcentration is required (typically, 1 min during CITE 3) followed by a 4-min period for separation and analysis. During CITE 3, the system was also configured to measure DMS (requires an oxidant scrubber). Since the oxidant scrubber interferes with the COS and CS$_2$ measurements, separate samples were collected for the DMS measurement. As a result of the sampling sequence of a DMS analysis followed by a COS/CS$_2$ analysis, separate COS/CS$_2$ measurements occurred about every 10 min. Precision of the measurements is estimated to be about 3% for mixing ratios in the range of 20–500 pptv. Accuracy (primary standard) was estimated to be 11%. Inflight calibrations (gas cylinder standard dynamically diluted) were performed at frequent intervals (e.g., one to two per hour). Additional discussion of the instrument is given in the references [Johnson and Lovelock, 1988; Johnson and Bates, this issue].

**TABLE 1. Sulfur Instrument Summary**

| Technique | Sulfur Measurements* | Sample Schedule† | Detection Limit, ‡ pptv | Precision, § | Accuracy, †% |
|-----------|----------------------|------------------|-------------------------|--------------|--------------|
| GC | COS | 3 min sample every 1/2 min | 10 | 2 | 10 |
| GC | H$_2$S | 3 min sample every 1/2 min | 25 | 5 | 25 |
| GC | SO$_2$ | 3 min sample every 1/2 min | 25 | 5 | 25 |
| GC | CS$_2$ | 3 min sample every 1/2 min | 25 | 5 | 25 |
| GC | DMS | 1 min sample every 10 min | 11 | 3 | 3 |
| MS | COS | 20 to 60 min collections | 14 | 4 | 14 |
| FLOR | H$_2$S | 20 to 60 min collections | 10 | 2 | 10 |
| Natusch | H$_2$S | 20 to 60 min collections | <1 | 2 | <1 |

*All gases not necessarily measured simultaneously.
†Typical sample schedule for CITE 3.
‡As stated at operations during CITE 3.
§Precision at 50 pptv.

Natusch (H$_2$S). The two applications of the Natusch technique differed only in minor procedural details (e.g., sample collection time, calibration of the fluorometer). Air is drawn through silver nitrate impregnated filters (Whatman 41) which are held in Teflon filter holders. Collection times range from 10 to 60 min, depending upon the expected H$_2$S mixing ratio (typically 10–20 min for boundary layer samples, 30–60 min for free-tropospheric samples). After collection, the filters are stored for postflight analyses (within 24 hours for CITE 3). For the analysis, the filters are washed with an NaOH/NaCN solution. The resulting sulfide concentration in the wash solution is quantified via a fluorescence quenching measurement (addition of mercuric acetate) using a calibrated fluorometer. To correct for interferences from atmospheric COS, two filters are used in tandem, with the...
second filter being used to measure the COS artifact. Results from the second filter are subtracted from the first filter to yield the H2S measurement value. Both investigators collected duplicate samples. The Max Planck Institute investigator team reported the results of duplicate sample independently, while the University of Miami investigators reported the average (normal method) of two duplicate samples. Inflight calibrations were not performed for either Natusch technique. The precision of the measurement is of the order of 14% (Max Planck Institute and for a 20–500 pptv range) and 10 and 7% (University of Miami) for mixing ratios of 20–100 and 100–500 pptv, respectively. The reader is referred to the references for detailed discussion of the techniques and minor differences in the procedures employed by the two investigators [Andreae et al., this issue; Bingemer et al., 1992; Cooper and Saltzman, this issue].

Data Protocol

The intercomparisons included gas-standard tests (ground) and inflight intercomparisons. Gas standards were prepared on-site at Wallops Island, Virginia, by personnel from the National Institute of Standards and Technology (NIST). The data protocol for the intercomparison of the standards and flight data was similar to that used for the other CITE 3 intercomparison species. Measurements were conducted blind with no exchange of information between the investigating teams prior to submittal of their results. Generally, final standards results, from the investigators and NIST, were submitted to the GTE project office within 48 hours after each test. Preliminary results from the airborne measurements were also submitted to the project office during the field operations. The results from the standards and flight measurements were analyzed by project personnel to monitor the progress of the tests and to provide input for subsequent tests.

As part of the data protocol for the ground standards test, each investigator or NIST had the option to declare a test invalid (when submitting the data) and to request a retest. For the flight intercomparisons, data protocol required all measurements taken to be reported. Along with the submitted data, the investigating teams provided a comment code as to the quality of the data. After submittal of all standards results for a gas, the results for that gas were discussed with the investigators during the field activities. Only a qualitative assessment of the progress of the flight intercomparison tests was provided to the investigators while in the field. Final flight data were submitted to the project within 3 months after completion of the field missions. These data were not normalized based on the results from the standards test. Detailed results of the flight intercomparisons (first release of results to the sulfur investigators) were discussed during a data workshop convened approximately 6 months after the field mission. After the workshop, only minor changes to the data base were made. None of these changes were significant in affecting the intercomparison results. Data changes made after the workshop are noted below.

Permission estimates of lower detection limits for many of the instruments were based upon laboratory results. Accordingly, all investigators were given an opportunity to reevaluate (based on workshop discussions) the lower detection limits for the techniques. Most investigators revised (lowered slightly) the detection limits. The data of Table I reflect the revised values. After the workshop, data from the FLUOR technique and for four flights from Wallops Island were resubmitted. Resubmitted values were 5.6% higher than previous values. The change in the resubmitted values was due to the use of an incorrect value for the inflight COS calibration standard in the initial analyses. In addition, one COS value, incorrectly submitted as valid data (flight 4 at 65 pptv), was resubmitted as calibration data. (It is noted that for flight 4, the original data file had numerous data at values of about 65 pptv which were properly identified as calibrations.) The results discussed reflect the resubmitted data.

Standards Intercomparison

The standards intercomparison was performed by having each instrument sample the output of a mobile reference source of COS, H2S, or CS2 provided by NIST. The output from the NIST system was sampled with the same sampling system (i.e., inlet, flow rates), as used during the aircraft flights. NIST values and uncertainties (about 10%) for the sulfur gas mixtures were based upon the value of the standards and subsequent dilution parameters.

Standards tests were conducted at Wallops Island, Virginia, from August 7 to 24, 1989. For H2S and CS2, a single mixing ratio (different for each investigator and in the range of 100–200 pptv) was provided to each instrument as installed aboard the aircraft. For the COS, two mixing ratios (range of 400–600 pptv) were provided. In general, and for a given instrument, the COS, H2S, and CS2 tests were performed on different days. Each investigator was required to provide at least three separate measurements of the NIST standard. The average of the values is compared to the NIST value to arrive at a level of agreement between NIST and the instrument, and the standard deviation on the average provides an estimate of the precision. The results of the standards tests are discussed later. A brief summary of test procedures and events which are important for the interpretation of the results is given below.

COS tests. COS tests were conducted per protocol with no retest or resubmittal of data.

H2S tests. During the H2S tests, one retest and two resubmittals of data occurred. In the first test of the GC technique, the investigator noted that the instrument was experiencing sensitivity problems and was not operating properly. The submitted data lacked precision, and results were inconsistent. Caveats submitted with the data were discussed (actual data not discussed) with the Science Team, and GTE project representatives suggested and scheduled a second test of the instrument. At the same meeting, results from the test of the Natusch technique (Max Planck Institute) were discussed. The investigator noted when the data were submitted and during discussions at the meeting that the submitted data may be low by 10% as the result of permeation tube damage during shipping (see also DMS companion paper [Gregory et al., this issue]). After discussion, it was decided that a retest was not required and that 10% was within the accuracy of the NIST supplied standards. At this meeting, all H2S ground-standards test results, except those of the GC technique, were released.

After release of the standards test results, the Max Planck Institute data were resubmitted (next day, new values 7% lower) due to an error discovered in data analyses associated with conversion to standard liters/minute (22.4 versus 24).
The resubmitted data improved the level of agreement with NIST from about 11 to 3%. Since the reason associated with the resubmitted data is not within the guidelines submitted with the original data (permeation tube damage) and the data were resubmitted after having released the \( \text{H}_2\text{S} \) results, both the original and resubmitted data were treated equally (i.e., the resubmitted data is not within the guidelines submitted with the original data). The resubmitted data improved the level of agreement with NIST, and data associated with the first test were resubmitted after having released the \( \text{H}_2\text{S} \) results, both with the original data (permeation tube damage) and the data associated with the first test. Since the data associated with the first test were resubmitted without the knowledge of the NIST, both the resubmitted and retest results are considered valid and are included in the analyses.

\( \text{CS}_2 \) tests

During the \( \text{CS}_2 \) standards tests, one retest and one resubmittal of data occurred, both associated with the GC technique. In discussion of the \( \text{CS}_2 \) standards results with the investigators (actual data not revealed), project personnel noted to the investigators that the gas chromato-graph–flame photometric technique test results (1) showed good precision, (2) did not show the same level of agreement with the NIST values as observed for the other two techniques, and (3) reported an absolute value for the NIST standard that was outside of the estimated accuracy and precision stated for the technique. The value submitted for the GC technique was 40% higher than the NIST value. (This information was not revealed in the discussions.) At the project’s recommendation, a retest of the GC technique was scheduled. Prior to the retest, the GC investigator resubmitted his test results for the first standards exchange. The resubmitted data were lower by 35% with an explanation that the initial \( \text{CS}_2 \) data submitted were based upon a \( \text{CS}_2 \) gas cylinder standard while the resubmitted data were based upon a perm tube standard. The investigator stated that the concentration of \( \text{CS}_2 \) in the gas cylinder standard was decaying with time. Results of the retest were consistent with the resubmitted results of the first test. Since the data for the first test were resubmitted without the knowledge of the magnitude and direction of the disagreement with NIST, both the resubmitted and retest results are considered valid and are included in the analyses.

**Aircraft Flights**

Twenty-one flights were conducted as part of the CITE 3 program. The first three were test flights based at the Wallops Flight Center (WFC), Virginia, and data obtained during these flights were designated “a priori” by the project as nonintercomparison data. Due to a logistical problem associated with operations in Natal, the last two flights (ferry from Brazil to Wallops Island) were also designated as nonintercomparison flights. The remaining 16 flights, including the ferry flights between WFC and Natal, were intercomparison data flights. As already noted, \( \text{COS} \), \( \text{H}_2\text{S} \), and \( \text{CS}_2 \) measurements were not made by all instruments during all flights. The measurements made by the various instruments are discussed later.

Flights were predominantly over the Atlantic Ocean off the coast of either the eastern United States or Brazil. Flights from WFC sampled the marine mixed layer and free troposphere at various distances from the continent. Natal flights were generally north and east from Brazil over the tropical Atlantic Ocean. Three night flights were flown from Natal. Flight altitudes ranged from 150 to 5000 m above sea level. Table 2 summarizes the 16 intercomparison flights. Details of the flights, type of air masses, and meteorological scenarios are discussed by Shipham et al. [this issue].

**DESCRIPTION OF ANALYSES APPROACH: FLIGHT DATA**

**Intercomparison Data**

As noted in Table 1, instrument sampling times and data reporting schedules were quite varied. As a result of this and the fact that each investigator routinely included calibration or maintenance periods in the sampling procedures, the CITE 3 intercomparisons were performed using structured sampling periods. To improve the temporal overlap of the various measurements, official intercomparison periods (IC periods) were designated for each flight. (IC periods were not used for the ferry flights 11 and 12.) IC periods were designed by considering the sampling schedules of all the sulfur instrumentation. Separate IC periods for each sulfur gas were not practical. Each IC period corresponded to a period of constant-altitude flight and was 30–60 min in duration. During the IC periods, each instrument followed a prescribed sampling schedule. Associated with each IC period was a preperiod or postperiod (5–15 min) designated
not be intercompared. In particular, an overlapped data base to identify outlier events during which measurements should be stated independently, and results, to identify data categories (i.e., subsets) under which measurements that were not representative of the overall data (time series) available for constructing overlapped data bases.

The intercomparison data used in the analyses are the data measured during the IC periods. A "simultaneous" or "overlapped" measurement is defined as having some overlap between any portion of the sample period reported by the investigators. The instrument/measurement having the longest integration time defined the overlap period, and as such, only a single measurement from that instrument is used for the overlap period. Where more than one value of COS, H₂S, or CS₂ is reported by any one of the remaining instruments during the defined overlap period, the arithmetic average of those measurements is used as the intercomparison value. Using this procedure, data bases were constructed by considering different combinations of measurement overlap (i.e., overlapping periods including data from all three techniques and combinations of two techniques). The term "data base" implies the ensemble of overlapped data periods constructed for a given combination of instruments and includes time periods from all 16 intercomparison flights for which an overlap occurred.

While the structured IC period approach tended to maximize the temporal overlap among the various measurements, overlap among the measurements was by no means always near 100%. For example and as illustrated in Figure 1 for H₂S (schematic of sampling schedule for the three techniques and a 50-min duration IC period), overlap between the two Natusch techniques was generally >90% (synchronized sample collection times), while overlap between a Natusch technique and the GC was seldom greater than 30%. While Figure 1 represents the H₂S data, it serves to illustrate the nature of the sulfur data (time series) available for constructing overlapped data bases and the wide range in the duration of measurement overlap represented in the data bases.

**Screening Analyses**

Each data set constructed was examined to evaluate measurements that were not representative of the overall results, to identify data categories (i.e., subsets) under which intercomparison results should be stated independently, and to identify outlier events during which measurements should not be intercompared. In particular, an overlapped data base was evaluated with the view to examine the influence of (1) the degree of temporal overlap (i.e., the ratio of common sample time of a measurement to the total duration of the overlap period), (2) data reported during periods in which significant ambient variations of the species was occurring, (3) the altitude at which the measurements were made, (4) systematic day-by-day variability, (5) the nature and type of air mass (total sulfur, water vapor, ozone, etc. content of the air), and (6) the mixing ratio value at which an overlapped data period occurred. In performing these analyses, numerous data correlations, regressions, confidence intervals, etc. were examined. Pertinent observations and conclusions from these analyses are presented as the data are discussed. Unless noted, the screening analyses identified no conditions for which agreement among the instruments differed from those discussed in the paper.

**DISCUSSION OF RESULTS**

**Standards Test Results**

While some of the biases between the instruments and NIST are statistically significant (95% confidence interval tests), it is concluded (when considering the uncertainty of the NIST gas mixtures and the measurements by the techniques) that all biases are within the uncertainty of the tests. A brief summary of the results is given below.

For COS, agreement between the measurements from the various instruments and the NIST values range from about -17% to +4% with a tendency for the MS (-17%) and FLUOR (-11%) techniques to be low. Both the MS and FLUOR biases are significant at 95% confidence levels. The associated precision (calculated as 2-sigma on the average of three to five samples) of the measurements were GC (6%) and 1% for the two mixing ratios tested), MS (2 and 1%), and FLUOR (4 and 2%).

For H₂S, agreement between the measurements from the instruments and the NIST values range from about -7% to +11%. As noted earlier, the 11% value is the Natusch-1 original data, which when resubmitted gave a +3% level of agreement with NIST. None of the biases are significant at a 95% confidence level.

For CS₂, agreement between the measurements and the NIST values range from about -8% to +14%. Again none of the biases are significant at a 95% confidence level.

If it is assumed that the NIST values contain no errors and that each instrument measurement is equally valid, then the standards data may be used to estimate an ensemble level of uncertainty that might be expected for a single COS, H₂S, or CS₂ measurement without reference to any one technique. The accuracy portion of the uncertainty is calculated from (1) as the average of the absolute bias. In the calculation the investigators' independent measurements (not the investigators' average) of the NIST standard are treated as equally valid and unbiased measurements of a known concentration (NIST). This is equivalent to assuming that from the COS standards test there are 22 replicate measurements (i.e., as compared to six averaged measurements) of a NIST standard.

\[
\text{Bias} = \frac{1}{N} \sum_{i=1}^{N} \frac{\text{NIST} - \text{instrument value}}{\text{NIST}}
\]  

(1)
The precision part of the uncertainty may be estimated from
the 1-sigma value on the average bias of (1) and a 95% confidence level calculation as

\[
\text{Precision} = \frac{t_{(\alpha/2, N-1)} \times (1 \text{ sigma})}{(N)^{1/2}}
\]  

where \( t \) is the Student's \( t \) statistic for \( \alpha = 0.5 \) and \( N - 1 \) degrees of freedom.

The accuracy portion from (1) is 12% for COS (\( N = 22 \) samples and 1 sigma of 6%), 7.6% for \( H_2S \) (12 samples and 1 sigma of 2.7%), and 8.2% for CS\(_2\) (12 samples and 1 sigma of 3.7%). The precision portion calculated from (2) is 2.7% (COS) and 2.4% (\( H_2S \) and CS\(_2\)). Thus, the expected ensemble uncertainty for the measurements (based solely on the standards test results and the assumption of instrument equality) is of the order of about 15, 10, and 11% for COS, \( H_2S \), and CS\(_2\), respectively. The \( H_2S \) uncertainty improves to about 8% if the original (first submittal, see earlier discussions) Natusch-1 data are excluded. Table 3 compares the stated accuracy and precision of the instruments (Table 1), the ground standards results, and the calculated uncertainties using (1) and (2).

**Carbonyl Sulfide Flight Intercomparisons**

**Typical flight data.** Figures 2 and 3 illustrate typical data obtained during the flights. The MS technique was configured to measure COS data only during ferry flights from Virginia to Brazil (Table 2, flights 11 and 12). The other two instruments provided intercomparison data during these same ferry flights, two flights from Virginia, and three flights from Brazil. Figure 2 is a time series of data from one of the ferry flights, and Figure 3 is from a night flight from Brazil. The A panels show the COS data, while the B panels are the time series of flight altitude, ozone, and dew point temperature. In general, while the COS data are plotted with horizontal bars to indicate the sample period for each sample, these bars are smaller than the plot symbol representing the data.

**Flight intercomparisons.** From data similar to those of Figures 2 and 3, intercomparison periods which involved some degree of sampling overlap between the various instruments were constructed. Two primary data bases resulted: (1) a data base (15 samples) which included overlapping data between the GC and the MS, and (2) a data base (36 samples) with overlapping data between the GC and FLUOR techniques. There were no time periods during which all three techniques had overlapping measurements, and only two periods which included overlapping data between the MS and FLUOR techniques. Measurements averaged over the longer time periods of an IC period are discussed later. Figure 4 represents the range of mixing ratios for the overlap periods of all COS data bases. For any given overlap period, each instrument sampled for about 30–50% of the total overlap period. The duration of an overlap periods are typically 3–8 min. Results from the screening analyses (see earlier discussion) showed no "abnormal" or nonrepresentative results; thus, no data are omitted or separated for special analyses. Figure 5, which shows the intercomparison results from the two primary data bases, is a plot of percent difference versus average COS mixing ratio. The definition of percent difference is given in the figure caption. Panel A is for the GC versus MS data base; panel B, the GC versus FLUOR data base. As indicated in the figure, agreement for single overlap periods was generally within about 10% with some indication that the GC might be high (same direction as the standards results) compared to the other two techniques. In a few cases (panel B), differences approached 20–25%. The average level of agreement (panels A and B) of about 5% is not significant at a 95% confidence level. Perhaps Figure 6, constructed as a box-and-whisker plot, best summarizes the results. The box-and-whisker plot offers the advantage of representing the range of agreement between the instruments in a single pictorial in which (1) the box encompasses 50% of the observations (box boundaries are the upper and lower interquartile ranges) and the horizontal line (within the box) notes the median of the data; (2) the whiskers (lines extending from the box) represent the extremes of the data or, in cases where there are values some distance for the bulk of the data, they extend to a value equal to 1.5 times the

---

**Table 3. Comparison of Instrument Accuracy/Precision With Ground Standards**

| Technique       | Stated Accuracy,* % | Stated Precision,* % | NIST Agreement,† % | Estimated Uncertainty,‡ % |
|-----------------|---------------------|----------------------|--------------------|---------------------------|
|                 |                     |                      | Average 2-sigma    | Accuracy | Precision | Total |
| Carbonyl Sulfide|                     |                      |                    |           |           |       |
| GC              | 20                   | 10                   | +2                 | 6.3       | 2.7       | 14.7  |
| GC-MS           | 20                   | 5                    | -17                | 1.4       |           |       |
| FLUOR           | 10                   | 3                    | -11                | 3.2       |           |       |
| Hydrogen Sulfide|                     |                      |                    |           |           |       |
| GC              | 20                   | 25                   | -7                 | 3.1       | 7.7       | 10    |
| Natusch-1       | 20                   | 14                   | +3                 | 9.4       |           |       |
| Natusch-2       | 20                   | 10                   | -6                 | 11.3      |           |       |
| Carbon Disulfide|                     |                      |                    |           |           |       |
| GC              | 20                   | 25                   | +5                 | 1.6       | 8.2       | 10.6  |
| GC-MS           | 20                   | 5                    | +14                | 1.7       |           |       |
| FLUOR           | 10                   | 3                    | -8                 | 4.9       |           |       |

*From Table 1.
†Results from ground standards test.
‡Calculated from (1) and (2) in the text.
Fig. 2. Time series of data for flight 11a from Virginia to Puerto Rico, September 9, 1989. Panel A is the time series of COS measurements from three instruments: gas chromatograph-flame photometric (GC), gas chromatograph-mass spectrometric (MS), and gas chromatograph/fluorination-electron capture (FLUOR). The sampling period for each measurement is shown as a horizontal bar (often smaller than the plot symbol). Panel B is corresponding time series for altitude, dew point temperature, and ozone.

Hydrogen Sulfide Flight Intercomparisons

Typical flight data. Figures 7 and 8 illustrate typical data obtained during the flights. With the exception of the GC technique during the ferry flights from Virginia to Brazil, all three techniques provided data on all flights. For the ferry flights, the GC was not configured for H$_2$S measurements in order to focus on COS. Figure 7 is a time series plot of data from one of the Virginia flights, and Figure 8 is the same night flight data from Brazil as the COS data of Figure 3. The horizontal bars shown with the Natusch symbols, which now extend beyond the size of the symbol, indicate the longer sampling periods associated with the measurements. The flags shown on the GC data (Figure 7) identify those values reported as at detection limits. All the GC data of Figure 8 were reported as such. As noted earlier, Natusch-1 took duplicate samples and reported both values.

Flight intercomparisons. Data bases which involved some degree of sampling overlap between the instruments were constructed from data similar to those of Figures 7 and 8. Duplicate samples of Natusch-1 were treated as separate measurements. The primary data base selected for discussion and which represents the results for all data bases is that which involved overlapping data from all three H$_2$S techniques. Figure 9 illustrates the range of mixing ratios for the data base. Of the 108 samples, eight overlapping data periods have been excluded from analyses (all from flight 9), as the submitted GC technique data were labeled as questionable. The screening analyses identified four overlapping data periods in which the GC instrument agreement with the
Natusch techniques appeared to be inconsistent with that observed from the total data base. All of these periods occurred during the Wallops flights and at times of suspected H$_2$S ambient variability. For these cases, temporal sampling overlap between the GC and Natusch techniques was only about 15%. The agreement between the two Natusch techniques for the same time periods was excellent as a result of nearly 100% sampling overlap. These four periods are also excluded from the 108-sample data base. The remaining 96-sample data base is the basis of the analyses presented.

As suggested in Figure 9, many of the overlapping data periods occurred at mixing ratios below the "estimated" 20 pptv detection limit of the GC technique (Table 1). In fact, of the 96 periods, 76 included data noted by the GC technique as an upper limit. Generally, the value reported was 10-15 pptv with a 1-sigma uncertainty of about 5 pptv. Those values submitted by the GC technique as upper limits or at detection limit are included in the intercomparison analyses at the value reported, e.g., data reported as at a detection limit of 10 pptv are assigned a value of 10 pptv in the analyses. (Intercomparison results are similar if these 76 samples are excluded and the remaining 20-sample data set is subjected to analyses.)

Figure 10 is a scatter plot of the 96-sample data base in which data from the GC and Natusch-2 are plotted against
data from Natusch-1 (X axis). The solid line represents the 1 = 1 line; the broken lines represent lines with slopes of 1.2 and 0.8. Linear regression results (assumes both X and Y are subject to error) are given in the figure. The values shown as plus or minus (±) are 1-sigma quantities. The 5-pptv offset associated with the GC regression reflects the large quantity of data that were reported as upper limits. Slope and intercept biases are significant at 95% confidence (ANOVA and 95% confidence interval tests).

Figure 11, another method of presenting the data of Figure 10, plots the delta difference between techniques (technique X minus technique Y) in a box-and-whisker format. As suggested by the figure, the level of agreement among the techniques is similar, agreeing 50% of the time to within about 5 pptv. The data shown by the symbols (termed outliers in box-and-whisker nomenclature) are mostly data at the higher mixing ratios. Figure 12 shows these delta values as a function of the average H₂S mixing ratio. The data of the figure are limited to averaged mixing ratios <50 pptv, where the average is calculated from values reported by all three techniques. Table 4 summarizes the delta analyses and shows the average delta (and ±1-sigma on the average) for

---

**Fig. 5.** COS intercomparison results. Panel A is the results from the data base constructed of overlapping measurements between the gas chromatograph-flame photometric (GC) and gas chromatograph-mass spectrometer (MS). Panel B is results from the data base involving the GC and gas chromatograph/fluorination-electron capture (FLUOR) techniques. Percent Difference is defined as [(GC) minus (MS)]/average COS (panel A) and [(GC) minus (FLUOR)]/average COS (panel B). Average COS is the average of the measurements reported for the overlap period from the respective two instruments. The average percent difference, 1 sigma on the average, and number of samples for each data base are given in the panels.
compare the relative agreement observed between the duplicate samples reported by Natusch-1.

From the data presented, one concludes that at the higher mixing ratios, instrument agreement is within about 15% (considering 95% confidence intervals for the slope biases of Figure 10) and at mixing ratios <25 pptv, agreement is generally within a few pptv (Table 4). Analyses were repeated using the official IC periods as overlap periods, and results were similar. For example, for mixing ratios <25 pptv (approximately 50 samples), delta differences (1 sigma) for GC minus Natusch-1, GC minus Natusch-2, and Natusch-1 minus Natusch-2 data bases are 2.5 (0.7), 2.9 (1.3), and 0.6 (1.0) pptv, respectively.

Carbon Disulfide Flight Intercomparisons

Typical flight data. Figures 13 and 14 illustrate typical data obtained during the flights. Flight 7 (Wallops) was the only flight in which all three techniques (GC, MS, and

![Fig. 6. Box-and-whisker plot of the data of Figure 5. The box-and-whisker format is discussed in the text.](image)

three groupings of the data: all mixing ratios, mixing ratios >25 pptv, and mixing ratios <25 pptv. Statistically, all delta biases of the table are significant at 95% confidence (confidence interval tests). The last two rows of data in Table 4
Fig. 8. Time series of data for flight 15, night flight from Brazil of September 16, 1989. Panel A is the time series of H₂S measurements from three instruments: gas chromatograph-flame photometric (GC) and two Natusch applications (Natusch-1 and Natusch-2). The sampling period for each measurement is shown as a horizontal bar (smaller than the plot symbol for the GC). All the GC data are measurements reported as an upper limit. Panel B is corresponding time series for altitude, dew point temperature, and ozone.

Flagged data points indicate FLUOR data reported as at the detection limit. The GC technique focused on measuring SO₂ during the Brazilian deployment and did not measure CS₂. In addition, while the FLUOR technique was configured to measure CS₂ for the Brazilian flights, all measurements reported were below its 2-pptv detection limit. During the Brazilian deployment, only the MS reported CS₂ mixing ratios as above detection limit. This is illustrated in Figure 14, which shows CS₂ data for the same night flight as the COS data of Figure 3 and the H₂S data of Figure 7. As was the case for the COS data, the horizontal bars representing the duration of the sampling periods are smaller than the size of the plot symbols used in Figures 13 and 14.

Flight intercomparisons. Data bases which involved some degree of sampling overlap between the various instruments were constructed. For the CS₂ data bases, temporal sampling overlap between the instruments was about 25% (data bases involving the FLUOR technique) and 75% (the GC versus MS data base). The higher temporal overlap for GC versus MS data base is the result of the similar sampling schedules used by the instruments (see Table 1). The screening analyses identified one overlap period which produced results which appeared to be inconsistent with the overall CS₂ intercomparison results; these data are excluded from the analyses and are flight 6 measurements in the continental boundary layer (off the U.S. coast) at about 150-m altitude. CS₂ values reported during this overlap period were 42.4 pptv (GC) and 2.3 pptv (FLUOR). Common sample time between the two measurements during the data period was only about 20%.

As already noted in Figure 13, flight 7 was the only flight in which all three techniques measured CS₂. Figure 15 represents the range of mixing ratios included in the 13 overlapping data periods constructed from flight 7, which included measurements from all three instruments. The
Fig. 9. Histogram representing the range of mixing ratios over which the $H_2S$ intercomparisons were performed.

Maximum range of the data is about 40 pptv with a large percentage of the intercomparisons limited to mixing ratios <10 pptv. Six of the 13 overlap periods of flight 7 included data reported by the FLUOR technique as below a detection limit of 2 pptv. These data are assigned a value of 2 pptv in the analyses.

Figure 16 plots the GC and FLUOR measurements against the MS measurements for the overlapping data periods of flight 7. Results from linear regression analyses (assumes that both the $X$ and $Y$ axis parameters are subject to error)

| REGRESSION RESULTS WITH NATUSCH-1 |
|-----------------------------------|
| Slope $= 1.16 \pm 0.06$ |
| Intercept $= -1.60 \pm 0.90$ pptv |
| $N = 96$ Samples |
| $r = 0.95$ |
| GC |
| Slope $= 0.80 \pm 0.05$ |
| Intercept $= 5.50 \pm 0.80$ pptv |
| $N = 96$ Samples |
| $r = 0.93$ |

are given in the figure. Slope and intercept biases noted by asterisks are statistically significant based upon 95% confidence interval testing. For the data of Figure 16, only the intercept biases are statistically significant with a tendency for the GC to be about 3 pptv high relative to the MS technique and the FLUOR technique to be about 3 pptv low compared to the MS. These intercept biases are similar to the delta differences calculated for the same data and given in Table 5 (all-overlap column). The last column of Table 5 shows that results are not different if the less-than data are excluded from the analyses. Restricting the analyses to mixing ratios <10 pptv and using data from the three data bases constructed by considering overlapping measurements between pairs of instruments, also gives results similar to those of Table 5. A closer inspection of these three data bases provides more information on $CS_2$ instrument agreement.

Figure 17 compares the GC and FLUOR measurements

Fig. 10. $H_2S$ scatter plot for overlapping data periods for data base involving overlapping measurements from all three instruments. Linear regression results (both $X$ and $Y$ subject to error) are given as slope, intercept, number of samples ($N$), and correlation coefficient ($r$). Values shown as plus or minus ($\pm$) are 1 sigma on the slope and intercept. All the slope and intercept biases are statistically significant at 95% confidence (confidence interval tests).

Fig. 11. Box-and-whisker plot of the data base of Figure 10 in which the delta difference between measurements is represented. The box-and-whisker format is discussed in the text.

Fig. 12. Delta difference plot of the box-and-whisker $H_2S$ data of Figure 11. Only data in which the average $H_2S$ is <50 pptv are plotted. Average $H_2S$ is the average of the GC, Natusch-1, and Natusch-2 mixing ratios reported or calculated for each overlap period.
TABLE 4. Delta Difference Results for H$_2$S

|                  | All     | <25 pptv | >25 pptv |
|------------------|---------|----------|----------|
| (GC) – (Natusch-1) | 3.0 ± 0.8 | 4.1 ± 3.3 | 2.7 ± 0.6 |
| (GC – (Natusch-2) | 2.1 ± 1.0 | −5.1 ± 3.4 | 4.0 ± 0.6 |
| (Natusch-1) – (Natusch-2) | −0.9 ± 0.8 | −9.2 ± 2.9 | 1.3 ± 0.4 |
| Number of samples | 96      | 20       | 76       |
| Duplicate Natusch-1 samples | 4.8 ± 9.3 | 12.5 ± 16.2 | 2.3 ± 1.9 |
| Number of samples | 61      | 15       | 46       |

The plus-or-minus (±) values are 1-sigma on the averaged delta difference. All delta differences are statistically significant at 95% confidence, i.e., 95% confidence interval does not include a delta of zero.

obtained during the only two flights in which both techniques measured CS$_2$. As suggested by the regression parameters (FLUOR values reported as at detection limits are plotted, but are excluded from regression analyses), instrument agreement is different between the two instruments for flights 6 and 7. Figure 18, panel A compares the delta difference between the GC and MS measurements as observed from the only two flights in which these two tech-
All FLUOR data are reported at lower detection limits

Fig. 14. Time series of data for flight 15, night flight from Brazil of September 16, 1989. Panel A is the time series of CS$_2$ measurements from the gas chromatograph-mass spectrometer (MS) and gas chromatograph/fluorination-electron capture (FLUOR) techniques. The sampling period for each measurement is shown as a horizontal bar (often smaller than the plot symbol). All the FLUOR data are measurements reported as at detection limit of 2 pptv. Panel B is corresponding time series for altitude, dew point temperature, and ozone.

Fig. 15. Histogram representing the range of mixing ratios over which the CS$_2$ intercomparisons were performed.
the form of a variable offset bias of a few pptv to perhaps as large as 6 or 7 pptv. A probable cause for a variable FLUOR bias may reside in the uncertainty associated with determining the measurement "blank" for a CS$_2$ measurement. As discussed by Johnson and Bates [this issue], the reproducibility of the system blank is the determining factor in establishing the lower detection limit (i.e., 2 pptv) for the CS$_2$ measurement. For CITE 3 the value for the system blank used in reporting the CS$_2$ data was 10 pptv [Johnson and Bates, this issue].

For all data bases, repeating the CS$_2$ analyses using the official IC periods to define the overlapping measurement periods gave similar results.

![Fig. 16. CS$_2$ scatter plot for overlapping data periods from flight 7 and the data base involving overlapping measurements from all three instruments, i.e., gas chromatograph-flame photometric (GC), gas chromatograph-mass spectrometer (MS), and gas chromatograph/fluorination-electron capture (FLUOR). Linear regression results (both X and Y subject to error) are given as slope, intercept, number of samples (N), and correlation coefficient (r). Values shown as plus or minus (+) are 1 sigma on the slope and intercept. Asterisks identify slope or intercepts that are statistically significant at 95% confidence (confidence interval tests).]

![Fig. 17. CS$_2$ scatter plots for the only two flights in which both the gas chromatograph-flame photometric (GC) and the gas chromatograph/fluorination-electron capture (FLUOR) techniques measured CS$_2$. Panel A is data from flight 6; panel B, flight 7. Regression parameters (both X and Y subject to error) are given in the panels as slope, intercept, number of samples (N), and correlation coefficient (r). The plus or minus (+) values are 1 sigma on the slope and intercept. Slopes and intercepts shown with an asterisk are statistically significant at 95% confidence based on confidence interval tests.]

### Table 5. Summary of Delta Analyses for CS$_2$: Flight 7

| Average Delta ± 1-sigma, pptv | All Overlap Periods | Excludes Overlaps With Less-Than Data |
|-----------------------------|---------------------|---------------------------------------|
| (Gas chromatograph) - (mass spectrometer) | 4.0 ± 4.0* | 3.4 ± 5.3 |
| (Gas chromatograph) - (fluorination) | 7.4 ± 4.0* | 6.9 ± 4.5* |
| (Mass spectrometer) - (fluorination) | 3.4 ± 2.6* | 3.5 ± 3.4* |
| Number of samples           | 13                  | 7                                      |

*Average delta differences which are statistically significant based on 95% confidence interval tests; i.e., 95% confidence interval does not include a delta of zero.
Fig. 18. Box-and-whisker plots comparing the delta difference between pairs of CS$_2$ instruments. Panel A compares the delta difference between the gas chromatograph–flame photometric (GC) and gas chromatograph–mass spectrometer (MS), i.e., delta = GC − MS, for flights 7 and 10, the only two flights in which both instruments measured CS$_2$. Panel B compares the delta difference between the gas chromatograph–mass spectrometer (MS) and gas chromatograph/fluorination–electron capture (FLUOR), i.e., delta = MS − FLUOR, for flights 7 and 8, the only two flights in which both instruments measured CS$_2$ above detection limits.

The flight data discussed above to obtain an estimate of the "ensemble" uncertainty associated with an aircraft measurement of the respective sulfur gases. In the analyses, it is assumed that each instrument provides an equally valid sulfur gas measurement during an overlapping data period, and that the calculated average from all values reported for a gas during the overlap periods is the "true" ambient mixing ratio of that gas (analogous to the true NIST value used in the standards calculations). These assumptions are reasonable based upon the CITE 3 data. While the analyses show that some biases among the instruments are statistically significant, none of the biases are of a magnitude or nature to suggest that any one instrument is providing invalid ambient measurements. Thus the average of the measurements from the various techniques during an overlap period is as accurate an estimate of the true ambient concentration as is available.

Applying these assumptions and the rationale developed in the discussion of (1) and (2) is equivalent to stating (using H$_2$S as an example) that there are three equally valid and independent H$_2$S measurements during each of the 96 overlap periods of the prime data base, and that the average of the H$_2$S measurement for each overlap is the true ambient H$_2$S concentration, i.e., 288 (3 times 96) measurements of a known (no error) H$_2$S mixing ratio. In those cases where there are multiple data bases, the data bases are combined and treated equally. For example, for COS, one data base had 15 overlap periods, the second data base had 36 samples, and each data base had two COS measurements for each overlap; thus there are 102 independent COS measurements (2 × 15 plus 2 × 36) of a known COS mixing ratio (average in each case).

Tables 6, 7, and 8 summarize the results for the COS, H$_2$S, and CS$_2$ ambient measurements, respectively. In Table 8 the data entry in the <10 pptv column only applies to the Wallops measurements, and the table footnote gives corresponding results for measurements during the Brazilian deployment. One interpretation of the data of the tables is that assuming the instruments are typical of those used to measure the various sulfur gases, then the indicated uncertainties represent the current confidence (without regard to the measurement technique) that can be attached to any reported ambient measurement of COS, H$_2$S, and CS$_2$.

Comparing these total uncertainties with similar values calculated from the standards test (Table 3), it is noted that the COS uncertainty of Table 6 is about a factor of 3 better than observed during the standards test, while the H$_2$S and CS$_2$ results (only the higher mixing ratio comparison is possible) of Tables 7 and 8 are about a factor of 2 poorer. Part of the explanation in the observed difference in the estimated uncertainty of the H$_2$S and CS$_2$ aircraft and ground standards results is attributed to the aircraft data

### TABLE 7. Uncertainty Estimate for H$_2$S Aircraft Measurements

| Mixing Ratio | Mixing Ratio |
|--------------|--------------|
| <25 pptv     | 228          |
| >25 pptv     | 60           |
| Number of independent samples | 228          |
| Accuracy portion (equation (1)), pptv | 2.4          |
| Precision portion (equation (2)), pptv | 0.3          |
| Total uncertainty, pptv | 2.7          |
| Percentual uncertainty | 15.0%          |

Data are for the 0–100 pptv mixing ratio range.

### TABLE 8. Uncertainty Estimate for CS$_2$ Aircraft Measurements

| Mixing Ratio | Mixing Ratio |
|--------------|--------------|
| <10 pptv     | 64           |
| >10 pptv     | 36           |
| Number of independent samples | 64           |
| Accuracy portion (equation (1)), pptv | 3.1   |
| Precision portion (equation (2)), pptv | 0.7   |
| Total uncertainty, pptv | 3.8          |
| Percentual uncertainty | 17.9%          |

Data are for the 0–50 pptv mixing ratio range.

*Includes only the data at Wallops. If same calculation is made for the data obtained during the Brazilian deployment (mass spectrometer and fluorination techniques), the accuracy, precision, and total uncertainty are 0.58, 0.15, and 0.73 pptv, respectively (66 samples).
having been obtained at mixing ratios generally a factor of 2–5 lower than the standards test data.

CONCLUSIONS

The CITE 3 intercomparisons provide data bases from which to evaluate COS, H$_2$S, and CS$_2$ measurements from an aircraft platform. For each gas, data from three techniques (only two fundamentally different detection principles for H$_2$S) are intercompared at various altitudes (0.2–5 km) in a predominantly marine environment. The COS intercomparisons occurred in a mixing ratio range of 400–600 pptv; the majority of the H$_2$S and CS$_2$ intercomparisons were at mixing ratios below 25 pptv and 10 pptv, respectively, with a maximum mixing ratio of about 100 pptv and 50 pptv, respectively. To intercompare the measurements from the techniques, data bases were constructed from time periods of simultaneous (overlapping) measurements among the techniques. For COS and CS$_2$, the total duration of an overlapping data period was typically 3–8 min, while for H$_2$S, overlapping data periods were 20–40 min in duration. While the data and discussions are limited to a few “selective” data bases, results are representative of all overlapping data bases analyzed.

The COS (carbonyl sulfide) results showed that individual measurements from the gas chromatograph-flame photometric, gas chromatograph-mass spectrometer, and gas chromatograph/fluorination-electron capture techniques generally agreed to within 10%; on the average to about 5%. There is a tendency (not statistically significant at 95% confidence) for the GC measurement to be high compared to the other two techniques (same trend observed in the ground-level standards tests). One concludes from the intercomparison data, the stated accuracy and precision for the techniques and available standards, and the analyses performed, that measurements from the techniques may be equally valid in terms of measuring true ambient levels of COS. Therefore the CITE 3 data are used to provide an estimate of the uncertainty associated with an airborne COS measurement without reference to the technique used. These analyses suggest that the total uncertainty associated with an ambient COS measurement is about 4%. A similar estimate made from the ground-level standards tests of the instruments installed on the aircraft (more controlled sampling conditions) gave a higher uncertainty of 15%.

The H$_2$S (hydrogen sulfide) results showed that measurements from the GC technique and the Natusch techniques (two separate applications of Natusch) agreed on the average to about 15%. For mixing ratios <25 pptv, agreement averaged about 5 pptv. While the observed biases were statistically significant (95% confidence interval testing), one concludes, based on the stated accuracy and precision for the techniques and available standards, that measurements reported by the techniques may be equally valid ambient measurements of H$_2$S. Using the CITE 3 data to provide an estimate of the uncertainty associated with an airborne H$_2$S measurement results in an uncertainty of about 18% (3 pptv for mixing ratios below about 25 pptv). Similar estimates using results from the ground-level standards tests gave an uncertainty of about 10%.

The CS$_2$ (carbon disulfide) results showed that only the MS had adequate sensitivity to measure the low CS$_2$ mixing ratios observed in clean, remote regions sampled over the Atlantic Ocean east of Brazil. Mixing ratios reported by the MS system for these regions were generally 1–2 pptv with an estimated detection limit of about 0.2 pptv. During these time periods the FLUOR technique reported CS$_2$ values as below a detection limit of 2 pptv. The GC technique did not measure CS$_2$ during the Brazilian deployment as a result of trade-off (3- to 5-min sample period) between detection limits for the CS$_2$ measurement and participation in CITE 3 intercomparison measurements for the other sulfur gases. Thus the CITE 3 data base does not provide an intercomparison of CS$_2$ measurements at these levels.

However, for the CS$_2$ data reported as above detection limit (few pptv to about 50 pptv), CITE 3 results showed that the GC, MS, and FLUOR techniques agreed on average to within a few pptv to about 6 or 7 pptv. A constant offset bias of about 4 pptv was observed between the GC and the MS techniques (flame photometric high). The offset bias between the FLUOR technique and the other two techniques was observed to be variable between flights and ranged from a few pptv to perhaps 6 or 7 pptv. While some of the observed biases were statistically significant based upon 95% confidence interval testing and suggest that CS$_2$ values reported by the various instruments are different, one concludes, based on the accuracy of the CS$_2$ standards and the stated accuracy for each technique, that measurements from any of the techniques in this range of mixing ratios can be considered to be a valid measurement of CS$_2$. Using the CITE 3 data to provide an estimate of the uncertainty associated with an airborne CS$_2$ measurement results in an uncertainty of about 20% (or 3 pptv) for mixing ratios between 10 and 50 pptv. A similar calculation for mixing ratios <10 pptv and using only the data measured during the Wallops deployment (data from all three techniques) gives about 4 pptv. A similar estimate using results from the ground-level standards tests (at about 100 pptv) gave an uncertainty of about 11%.

Acknowledgments. The authors would like to acknowledge all CITE 3 participants and express thanks for the cooperation, support, and professionalism displayed throughout CITE 3. In particular, we acknowledge Linda Warren and the valuable contribution made in developing the numerous software packages required to construct the intercomparison data bases. Special recognition is extended to the Wallops Electra flight and ground crews for the long hours spent in the field preparing and flying the aircraft while based at Wallops Island, Virginia, and Natal, Brazil. To the hangar crew at Wallops, their support and helpful suggestions during installation of equipment and test flights at Wallops are appreciated. Lastly, we acknowledge the support and hospitality of the government and citizens of Brazil during the Natal portion of the field program, particularly Volker W. J. H. Kirchhoff and Adauto Gouvia Motta of the Instituto de Pesquisas Espaciais.

REFERENCES

Andreae, M. O., The emission of sulfur to the remote atmosphere, in The Biogeochemical Cycles of Sulfur and Nitrogen in the Remote Atmosphere, edited by J. N. Galloway et al., pp. 5–25, D. Reidel, Hingham, Mass., 1985.

Andreae, M. O., Ocean-atmosphere interactions in the global biogeochemical sulfur cycle, Mar. Chem., 30, 1–29, 1990.

Andreae, M. O., and W. A. Jaeschke, Exchange of sulfur between biosphere and atmosphere over temperate and tropical regions, in Sulfur Cycling in Terrestrial Systems and Wetlands, edited by R. W. Howarth and J. W. B. Stewart, John Wiley, New York, in press, 1993.

Andreae, T. W., M. O. Andreae, H. G. Bingemer, and C. Leck,
Gregory, G. L., et al., An intercomparison of airborne nitric acid measurements, J. Geophys. Res., 95, 10,089–10,102, 1990d.

Gregory, G. L., L. S. Warren, D. D. Davis, M. O. Andreae, A. R. Bandy, R. J. Ferek, J. E. Johnson, E. S. Saltzman, and D. J. Cooper, An intercomparison of instrumentation for tropospheric measurements of dimethyl sulfide: Aircraft results for concentrations at the parts-per-trillion level, J. Geophys. Res., this issue.

Hoell, M. J., et al., An intercomparison of airborne monoxide, nitric oxide, and hydroxyl measurement techniques: Overview of results, J. Geophys. Res., 89, 11,819–11,825, 1984.

Hoell, J. M., G. L. Gregory, D. S. McDougal, M. A. Carroll, M. McFarland, B. A. Ridley, D. D. Davis, J. Bradshaw, M. O. Rodgers, and A. L. Torres, An intercomparison of nitric oxide measurement techniques, J. Geophys. Res., 90, 12,843–12,851, 1985a.

Hoell, J. M., G. L. Gregory, D. S. McDougal, G. W. Sachse, G. F. Hill, E. P. Condon, and R. A. Rasmussen, An intercomparison of carbon monoxide measurement techniques, J. Geophys. Res., 90, 12,881–12,889, 1985b.

Hoell, M. J., et al., Operational overview of the NASA GTE/CITE 3 airborne instrument intercomparisons for sulfur dioxide, hydrogen sulfide, carbonyl sulfide, dimethyl sulfide, and carbon disulfide, J. Geophys. Res., this issue.

Johnson, J. E., T. S. Bates, Atmospheric measurements of carbonyl sulfide, dimethyl sulfide, and carbon disulfide using the electron capture detector, J. Geophys. Res., this issue.

Johnson, J. E., and H. Harrison, Carbonyl sulfide concentrations in the surface waters and above the Pacific Ocean, J. Geophys. Res., 91, 7883–7888, 1986.

Johnson, J. E., and E. J. Lovelock, Electron capture sulfide detector: Reduced sulfur species detection at the femtomole level, Anal. Chem., 59, 238–242, 1977.

Lewin, E. E., M. Lalievic, and A. R. Bandy, Determination of atmospheric carbonyl sulfide by isotope dilution gas chromatography/mass spectrometry, Anal. Chem., 59, 1802–1805, 1987.

Maroulis, P. J., and A. R. Bandy, Measurements of atmospheric concentrations of CS$_2$ in the eastern United States, Geophys. Res. Lett., 7, 681, 1980.

Maroulis, P. J., A. L. Torres, and A. R. Bandy, Atmospheric concentration of carbonyl sulfide in the southeastern and western United States, Geophys. Res. Lett., 4, 510, 1977.

McFarland, B. A. Ridley, R. C. Harriss, and J. M. Hoell, NASA Global Tropospheric Experiment, EOS Trans. AGU, 64, 561–562, 1993.

National Aeronautics and Space Administration, Research needs in heterogeneous tropospheric chemistry, NASA Conf. Publ., 2320, 15–17, 1984.

Rasmussen, R. A., M. A. K. Khalil, and S. D. Hoyt, The oceanic source of carbonyl sulfide, Atmos. Environ., 16, 1591–1594, 1982.

Rodhe, H., and I. Isaksen, Global distribution of sulfur compounds in the troposphere estimated in a height/latitude transport model, J. Geophys. Res., 85, 7401–7409, 1980.

Saltzman, E. S., and D. J. Cooper, Shipboard measurements of atmospheric dimethyl sulfide and hydrogen sulfide in the Caribbean and Gulf of Mexico, J. Atmos. Chem., 7, 191–209, 1988.

Shipham, M. C., A. S. Bachmeier, and B. E. Anderson, CITE 3 meteorological highlights, J. Geophys. Res., this issue.

Sze, N. D., and M. F. Ko, CS$_2$ and COS in the stratospheric sulfur budget, Nature, 280, 308–310, 1979.

Toon, O. B., R. P. Turco, P. Hamill, C. S. Ciarn, and R. C. Whitten, A one-dimensional model describing aerosol formation and evolution in the stratosphere, 2, Sensitivity studies and comparison with observations, J. Atmos. Sci., 36, 718–736, 1979.

Toon, O. B., J. F. Kasting, R. P. Turco, and M. S. Liu, The sulfur cycle in the marine atmosphere, J. Geophys. Res., 92, 943–963, 1987.

Torres, A. L., P. J. Maroulis, A. B. Goldberg, and A. R. Bandy, Atmospheric OCS measurements on Project GAME TAG, J. Geophys. Res., 85, 7357–7360, 1980.

Bandy, A. R., J. E. Johnson, E. S. Saltzman, and D. J. Cooper, J. Geophys. Res., this issue.

Bandy, A. R., J. E. Johnson, E. S. Saltzman, and D. J. Cooper, J. Geophys. Res., this issue.

Bandy, A. R., P. J. Maroulis, L. Shalaby, and L. A. Wilner, J. Geophys. Res., 88, 3732–3740, 1983.

Berresheim, H., and W. Jaeschke, The contribution of volcanoes to the global atmospheric sulfur budget, J. Geophys. Res., 88, 1180, 1981.

Bingemer, H. G., S. Burgermeister, R. L. Zimmermann, and H. W. Georgii, Atmospheric OCS: Evidence for a contribution of anthropogenic sources, J. Geophys. Res., 95, 6015–6025, 1990.

Bingemer, H. G., M. O. Andreae, T. W. Andreass, P. Artaxo, G. Helas, D. J. Jacob, N. Mihalopolous, and B. C. Nguyen, sulfur gases and aerosols in and above the equatorial African rain forest, J. Geophys. Res., 97, 6207–6217, 1992.

Brasseur, G. P., C. Granier, and S. Walters, Future changes in stratospheric ozone and the role of heterogeneous chemistry, Nature, 348, 626–628, 1990.

Carroll, M. A., Measurements of COS and CS$_2$ in the free troposphere, J. Geophys. Res., 90, 10,483–10,486, 1985.

Charlson, R. J., J. E. Lovelock, M. O. Andreae, and S. G. Warren, Oceanic phytoplankton, atmospheric sulfur, cloud albedo, and climate, Nature, 326, 655–661, 1987.

Cooper, D. J., and E. S. Saltzman, Measurements of atmospheric dimethyl sulfide, hydrogen sulfide, and carbon disulfide during GTE/CITE 3, J. Geophys. Res., this issue.

Crutzen, P. J., The possible importance of COS for the sulfate layer of the stratosphere, Geophys. Res. Lett., 3, 73–77, 1976.

Cullis, C. F., and M. Hirschler, Atmospheric sulfur: Natural and man-made sources, Atmos. Environ., 14, 1263–1278, 1980.

Emmons, L. K., and P. J. Maroulis, A. L. Torres, and A. R. Bandy, Atmospheric measurements of dimethyl sulfide, hydrogen sulfide, and carbon disulfide during GTE/CITE 3, J. Geophys. Res., this issue.

Ferek, R. J., and E. S. Saltzman, Measurements of atmospheric dimethyl sulfide, hydrogen sulfide, and carbon disulfide using the SCOPE 24, edited by B. Bolin and R. Cook, pp. 56–61, John Wiley, New York, 1983.

Ferek, R. J., and E. S. Saltzman, Reduced sulfur species detection at the femtomole level, J. Geophys. Res., this issue.

Ferek, R. J., and M. O. Andreae, The supersaturation of carbonyl sulfide in surface waters of the Pacific Ocean off Peru, Geophys. Res. Lett., 10, 939–936, 1983.

Ferek, R. J., and M. O. Andreae, Photochemical production of carbonyl sulfide in marine surface waters, Nature, 307, 148–150, 1984.

Freneny, J. R., M. V. Ivanov, and H. Rodhe, The sulphur cycle, in The Major Biogeochemical Cycles and Their Interactions (SCOPE 24), edited by B. Bolin and R. Cook, pp. 56–61, John Wiley, New York, 1983.

Goldberg, A. B., P. J. Maroulis, L. A. Wilner, and A. R. Bandy, Studies of H$_2$S emissions from a salt water marsh, Atmos. Environ., 15, 11, 1981.

Gregory, G. L., J. M. Hoell, S. M. Beck, D. S. McDougal, J. A. Meyers, and D. B. Brunt, Operational overview of Wallops Island instrument intercomparison: Carbon monoxide, nitric oxide, and hydroxyl instrumentation, J. Geophys. Res., 90, 12,808–12,818, 1985.

Gregory, G. L., J. M. Hoell, B. A. Ridley, H. B. Singh, B. Gandrud, L. J. Salas, and J. Shetter, An intercomparison of airborne PAN measurements, J. Geophys. Res., 95, 10,077–10,088, 1990a.

Gregory, G. L., A. R. Bandy, R. J. Ferek, J. E. Johnson, E. S. Saltzman, and D. J. Cooper, An intercomparison of airborne nitrogen dioxide instruments, J. Geophys. Res., 95, 10,103–10,128, 1990b.

Gregory, G. L., J. M. Hoell, A. L. Torres, M. A. Carroll, B. A. Ridley, M. O. Rodgers, J. Bradshaw, and D. D. Davis, An intercomparison of airborne nitric oxide measurements: A second opportunity, J. Geophys. Res., 95, 10,129–10,138, 1990c.
Turco, R. P., R. C. Whitten, O. B. Toon, J. B. Pollack, and P. Hamill, OCS, stratospheric aerosols and climate, Nature, 283, 283-286, 1980.

Turner, S. M., and P. S. Liss, Measurements of various sulphur gases in a coastal marine environment, J. Atmos. Chem., 2, 223-232, 1985.

M. O. Andreae, Anteilung Biogeochemie, Max-Planck-Institut fur Chemie, Postfach 3060, D-6500 Mainz, Germany.

A. R. Bandy and D. C. Thornton, Chemistry Department, College of Science, Drexel University, Philadelphia, PA 19104.

J. D. Barrick and G. L. Gregory, Atmospheric Sciences Division, Mail Stop 483. NASA Langley Research Center, Hampton, VA 23681-0001.

D. D. Davis, School of Earth and Atmospheric Sciences, Room 108-B, Georgia Institute of Technology, 220 Third Street, Atlanta, GA 30332-0340.

J. W. Johnson, NOAA Pacific Marine Environmental Laboratory, 7600 Sand Point Way NE, Seattle, WA 98115.

E. S. Saltzman, Division of Marine and Atmospheric Chemistry, Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149.

(Received February 4, 1992; revised March 16, 1993; accepted March 16, 1993.)