In Situ Processing of Boundary Layer Flow Measurements For Data Compression[1]

Albert J. Williams 3rd
Applied Ocean Physics & Engineering Dept.,
Woods Hole Oceanographic Institution

and

Thomas F. Gross
Skidaway Institute of Oceanography

Introduction

Turbulent boundary layer flows are not fully characterized by the simple velocity means. Higher moment averages of the turbulent velocities such as Reynolds stress, turbulent kinetic energy and vector averaged flux of scalar constituents must be obtained. However, in long deployments, internally recording instruments cannot store the quantity of data required by these computations. *In situ* processing of the raw data time series to yield these statistics has proven to be the solution to data recording limitations.[1]

BASS, an acoustic current meter array, allows measurements of full velocity vectors at six levels through the turbulent boundary layer with a surprisingly low power consumption.[2] At a sampling rate of 2 Hz some 14 megabytes of data may be produced each day or about one gigabyte in a two and a half month experiment. By using an *in situ* processing scheme BASS recorded the 2 Hz data averaged to thirty minute means of velocity and vector flow component products for a seven month experiment with two megabytes of storage. Half-hour means of velocity and higher moments are sufficiently long to invoke stationarity for turbulent boundary layer eddies in the sea allowing the Reynolds decomposition to be used, further compressing the data to simple means and means of products of fluctuations. A large part of the information contained in the full 2 Hz data stream is therefore retained by the cross variance calculations allowing a huge reduction in storage requirements. However, spectral information is lost for periods shorter than one hour. To extend the spectrum of velocity to higher frequency, two minute averages of velocity from one sensor are included with the thirty minute means.

Episodic bursts of raw data are also recorded by BASS for one hour upon event trigger. Four such events can be recorded in a second two megabyte storage module. In addition to extending the spectrum of velocity to one hertz, structure in the flow is accessible from such measurements, acquired at times when changes are anticipated. Events have been triggered on successively higher suspended sediment concentrations, on current speed, and on turbulent kinetic energy. In a special free drifting configuration,[3,4] events were triggered by temperature variance.

At the time of an event, the spectrum of velocity, temperature, sediment concentration, and pressure can be obtained from 1 Hz to 0.003 Hz with five degrees of freedom. From the two minute means, the spectrum of velocity can be obtained from 0.004 Hz to $3.3*10^{-7}$ Hz with five degrees of freedom. The spectrum of stress, turbulent kinetic energy, and flux can be obtained from $2.8*10^{-4}$ Hz ($1\ Hr^{-1}$) to $3.3*10^{-7}$ Hz ($1.2*10^{-3}\ Hr^{-1}$).

Data Compression

Increased sampling frequency, the number of channels sampled, and the duration of unattended deployment increases vastly the total quantity of data collected (Fig. 1).

Increased storage capacity and data telemetry tend to reduce the stress of these vast quantities of data on future systems but neither can solve the problem without *in situ* data compression. Cassette recorders of two megabyte capacity were introduced in 1971, cartridge recorders of 60 megabyte capacity were first used in 1985, and optical discs of 120 megabyte capacity were used in 1988. In 1990 200 megabyte hard discs became available (ONSET, Inc. Tattletale Model VI). With each increment in storage capacity the demands and ambitions of the scientist increment faster. The half hour means of velocity and stress stretched the capability of the two megabyte storage media. Now that 200 megabytes are available there is a demand to obtain directional wave spectra information which can be recorded only every hour due to storage limitations of the 200 megabyte device. If one assumes the doubling time for storage capacity in underwater instruments is three years, it is unlikely that storage capacity can catch up with demand. Williams' experience serves as an example. In 1972, he sampled 4 channels to 16 bits at 5 Hz for four hours for a total output of half a megabyte. In 1978, he sampled 32 channels to 12 bits at 0.6 Hz for thirteen hours for a total output of 1.4 megabytes. In 1981, he was sampling 48 channels to 16 bits at 2 Hz for five hours for a total output of 3.5 megabytes. This exceeded the recorder capacity available at that time by almost a factor of two, but instead of adding a second recorder, he performed a routine subtraction *in situ* to halve the data rate. In 1985, Williams and Gross were sampling 48 channels to 16 bits at 2 Hz for four months for a total output of 2 gigabytes. No recorder of that capacity was available then either. The new 60 megabyte recorder was inadequate by a factor of 30, so the *in situ* processing scheme here discussed was employed. In 1988, we were sampling 60 channels to 16 bits at 6.7 Hz for two months for a total output of four gigabytes. In this case, two levels of *in situ* processing were used, correlations and FFTs.

Data telemetry at high baud rates over short distances and at low baud rates over satellite link for extended deployments is now available. But here too, increases in data acquisition rates stress available communication channels without *in situ* processing. A short range, short-lived VHF radio telemetry link Williams used in 1981 would handle 4800 baud. With 40 channels sampled to 16 bits at 5 Hz, the data was 400 bytes/sec. With a 12 bit/byte encoding for serial transmission, this matched the 4800 baud capacity of the channel. In 1986, the sample rate was increased to 25 Hz and *in situ* processing was needed to reduce this rate to the 4800 baud available. Telemetry of raw

[1]WHOI Contribution No. 7742

0-7803-0202-8/91/0000-0973$1.00©1991 IEEE

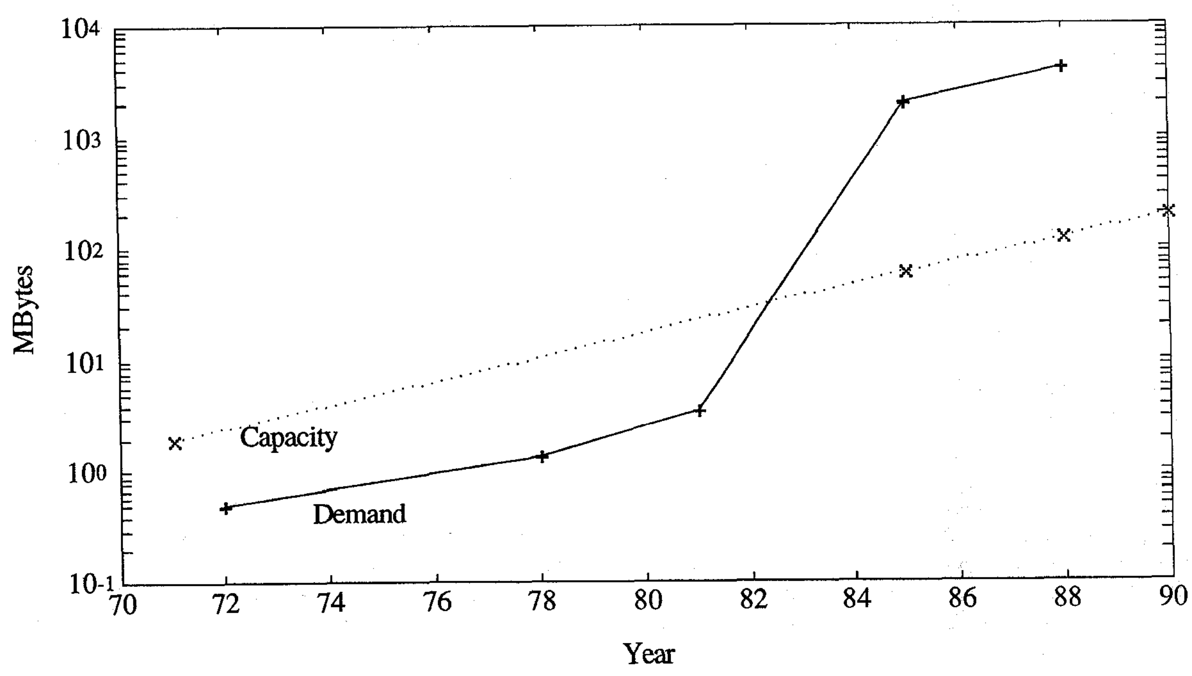

Figure 1. Williams' demand for data capacity vs. available underwater logger capacity. Demand outpaces supply. Data compression is required.

data is also limited by power considerations. A 5 watt transmitter working at 1200 baud will telemeter only 86 kilobytes per watt-hour. A large battery pack of 2 Kwh provides a total of only 170 megabytes, less than modern internal recorder capacity. Satellite links typically have a data throughput between 0.1 and 10 baud. Clearly, data compression is required here again.

Technique of Reynolds Decomposition

Turbulent flow may be considered to have a steady component and a fluctuating component. The steady component is the average over a time that is large compared to the longest period of fluctuation. In practice, it is rarely clear where turbulent fluctuations stop and mean flow variations start. However, for any selected interval, a Reynolds decomposition can be made into a mean and fluctuating part:

$$\mathbf{u} = \mathbf{U} + \mathbf{u}'$$

where $\mathbf{u}$ is the instantaneous vector velocity, $\mathbf{U}$ is the mean over the interval, and $\mathbf{u}'$ is the instantaneous velocity fluctuation.

For boundary layer studies, velocity means, Reynolds stress components, and turbulent kinetic energy components are instructive. Velocity means are simply $\mathbf{U}$, Reynolds stress and turbulent kinetic energy components are derived from the terms in the Reynolds stress tensor $\tau_{ij} = u_i' u_j'$.

The term $\tau_{ij} = \overline{u_i' u_j'}$, can be found by expanding the Reynolds decomposition and averaging.

$$\begin{aligned} \tau_{ij} &= \overline{(u_i - U_i)(u_j - U_j)} \\ &= \overline{u_i u_j} - \overline{u_i U_j} - \overline{u_j U_i} + \overline{U_i U_j} \end{aligned}$$

Bringing constants outside the averages

$$\begin{aligned} \tau_{ij} &= \overline{u_i - u_j} - \overline{u_i} U_j - \overline{u_j} U_i + U_i U_j \\ &= \overline{u_i u_j} - U_i U_j. \end{aligned}$$

The mean velocity components can be obtained by adding each vector component sample to the accumulated sum for the interval over which the computation will be made and dividing by the number of samples. The products are similarly formed as the samples are taken, accumulated over the interval used in the computation, and divided by the number of samples so accumulated. No large array need be acquired, only accumulators for the velocity sums, the product sums, and the counts for each accumulator. Thus the computations can be done as the measurements are made.

At the end of the period over which the means will be taken, a longer computation must be made; the means must be formed and the subtraction made. While this is trivial in principle, it is more complex in practice because some measurements may have been discarded as errors so the counts for each accumulation may be different. Since a count is retained for each accumulation, such a computation can be done. The significance in the product of the fluctuations is retained by subtracting the product of the means from the mean of the products. Care must be taken to avoid all round off errors as τ_{ij} will be less than one percent of the mean squared.

The 1802 microprocessor used in these *in situ* computations (a stodgy work horse by current standards) was capable of timing and reading the A/D conversions of 48 channels and performing the above accumulations at up to 25 Hz. Direct assembly language programming was necessary to achieve these rates. Recently faster microprocessors have been put into use, but the speed is usually used up by high level language interfaces. It is still necessary to write complicated *in situ* processing programs in assembly code.

The constraint of *in situ* processing of means and variances is loss of information about correlations and spectra that have not been saved. They cannot be reconstructed from the averages or the processed data. However, more than one process at a time can be applied to the data. Velocity can be correlated with scalar quantities to measure flux. The process is similar to that used for the Reynolds stress in which the product of the means is subtracted from the mean of the products. Correlation of temperature with velocity gives heat flux and correlation of sediment concentration with velocity gives sediment transport.

Spectral information at periods less than the averaging period is lost even though many correlations are performed and saved. Spectral analysis requires that the time series be temporarily saved. This puts a greater demand on working memory since 30 minutes of 1200 baud data is 216 kilobytes. The FFT (Fast Fourier Transform) method used to obtain spectra requires the entire data set for the block to be present at once. We added a microcomputer with 512 kilobytes of RAM in 1988 to perform FFTs on our velocity and pressure data. An integer FFT algorithm was used on each of ten variables - pressure, U and V from two sensors, W from five sensors - sampled at 1.67 Hz. For each variable, five ensembles of 512 samples each were saved to obtain a 256 element transform with ten degrees of freedom. At two bytes per sample, this consumed 50 kilobytes of memory, one tenth of that available but twenty-five times that used for the Reynolds stress correlation. Data were acquired for 25 minutes, 36 seconds out of the thirty minute recording interval to obtain these five blocks, and the 50 FFTs were performed on one sample while the next data set was being stored in a second 50 kilobyte buffer in ping-pong mode. The time to compute each FFT was 8 seconds, taking 6 minutes 40 seconds for the complete calculation.

A full Fourier transform retains as much information as the original signal. After averaging five spectra and keeping only the magnitude, a data reduction of only a factor of ten is achieved. Therefore, only part of the spectra can be recorded. The experimental purpose dictates what part of the spectra should be retained. As only surface gravity waves of periods 7 to 120 seconds can affect bottom boundary layer processes in 100 meter depths, just the lowest forty frequencies of the spectra were retained. This gave a data reduction of a factor of sixty over the original data. In 1988 the FFTs added 512 bytes to the velocity averages, Reynolds stresses and other miscellaneous sampled data for a total of 768 bytes each half hour. This allowed a little less than two months of data to be recorded on our existing tape units.

In 1990, our recording capacity was increased to 20 megabytes with a magnetic hard disk logger (Onset Tattletale VI) and the raw data for the last FFT was written to disk, along with the previous 20 FFTs, averages, and correlations, each ten hours. These raw samples will allow us to try other correlations and structural investigations of the flow. But the increased demand on memory, another factor of ten over the 1988 system, along with the small sample available for this exploratory analysis, encourages us to consider implementing additional correlations and FFTs of correlated quantities. Late in 1990, we added a 200 megabyte hard disk and temporarily dispensed with FFT processing. But only a fraction of the data were saved. Again it begs the question of how to use storage capacity. And even when capacity increases rapidly, data rates increase more rapidly. The spurt in capacity from 1988-1990 will be soaked up by greedy sensors and ambitious scientists in 1992.

If the 200 megabytes of available storage were divided into 100 bins each of which recorded a time series of some statistical quantity, half hourly measures could be saved for at least four months. Now it is up to the imagination and ingenuity of the user to specify in advance of the experiment what 100 statistical measures should be computed. It is easy to expand the present list of velocity means, velocity correlations, FFTs of velocity, FFTs of pressure, and spot samples of temperature, optical transmission, and attitude to 10 or 15 including heat flux and sediment transport. It is more difficult to add the next 15 or 20 including spatial correlations between sensors separated vertically. It is a real challenge to get to fifty statistical measures, a range exceeding anything customarily analyzed in boundary layer flow data sets. And there will still be another fifty available.

When devising an *in situ* processing scheme the scientist must always be aware that some information will be lost. While physical insight can guide one to select reasonably appropriate quantities, the unknown must not be assumed to be unimportant. Therefore in almost all of our experiments some of the raw data was retained. Post processing analysis of the raw data and comparison to the *in situ* calculations inevitably yields insight to guide the development of future *in situ* processing schemes. For instance the spectral calculations of pressure and velocity were found to be useful, but upon analysis of the full data stream it became apparent that the more complicated calculation of directional wave spectra would be necessary in future experiments. Figure 2 shows the spectrum of waves through a storm on day 356 of 1988. This is the spectrum of pressure. But higher level processing is required to obtain directional spectra. When designing an instrument which uses *in situ* processing methods, flexibility must be built in from the beginning.

In conclusion, means and correlations are an effective way to compress data by *in situ* processing with modest working memory requirements. FFTs retain spectral information at the cost of increased, but still modest working memory requirements. Long duration experiments and greedy sensors will outpace even the rapid increases in storage capacity. But smart *in situ* processing can accommodate these demands and even adapt high sample rates to low data transmission rates. The final constraint, bandwidth for data return from expendable instruments, may ultimately prove the most limiting. But clever *in situ* processing can accommodate this constraint as well.

Acknowledgments

The authors would like to thank Dr. J. Kravitz at ONR for his long-term support to the HEBBLE and STRESS programs which stimulated the demand for data compression. The efforts behind this paper were supported by ONR grant N00014-89-J-1058.

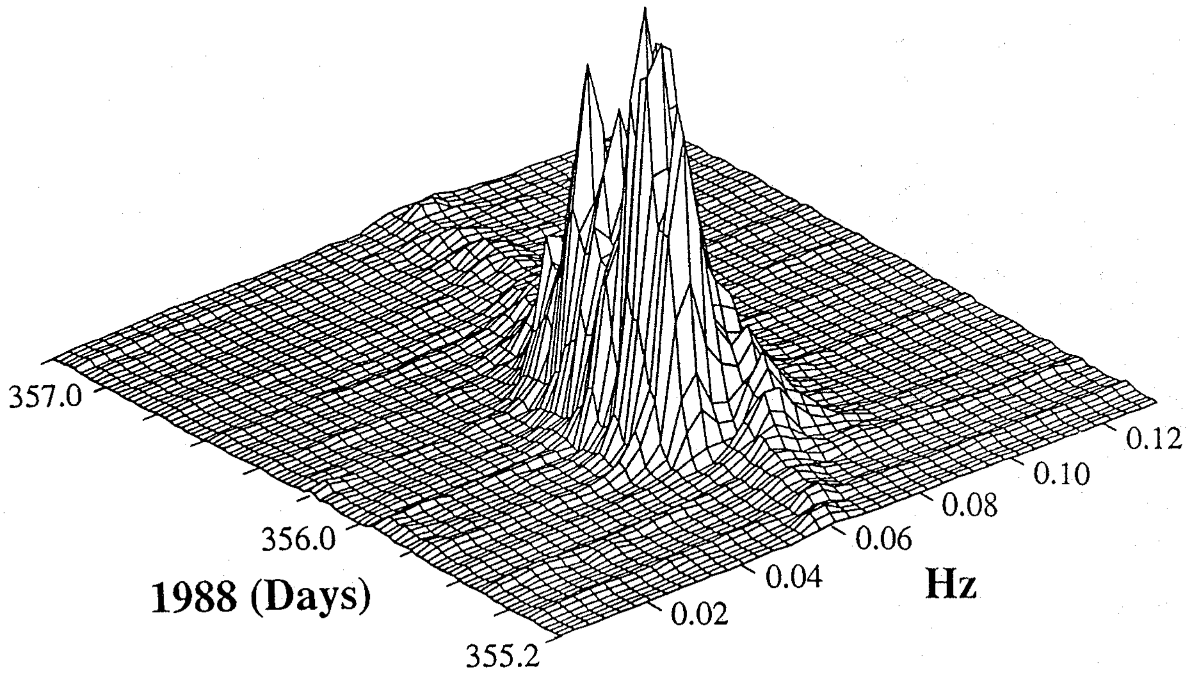

Figure 2. Spectrum of shelf floor pressure computed *in situ* from 1.67 Hz measurements transformed by integer FFT every 30 minutes.

References

1. T.F. Gross, A.J. Williams 3rd, and W.D. Grant, "Long-Term *In Situ* Calculations of Kinetic Energy and Reynclds Stress in a Deep Sea Boundary Layer," J. Geophys. Res., vol. 91(C7), pp. 8461-8469, July 1986.

2. A.J. Williams, 3rd, J.S. Tochko, R.L. Koehler, et al., "Measurement of Turbulence in the Oceanic Bottom Boundary Layer with an Acoustic Current Meter Array," J. Atmos. and Oceanic Technol., vol. 4(2), pp. 312-327, June 1987.

3. A.J. Williams, 3rd, C.H. Converse, and J.W. Nichoison, "Richardson Number Float," Current Practices and New Technology in Ocean Engineering, 1987-OED vol. 12, ed. G.K. Wolfe, The American Society of Mechanical Engineering.

4. E. Kunze, A.J. Williams III, and M.G. Briscoe, "Observations of shear and vertical stability from a neutrally buoyant float," J. Geophys. Res., vol. 95(C10), pp. 18,127-18,142, October 1990.