Journal of Marine Research, 43, 1–12, 1985

Journal of MARINE RESEARCH

Volume 43, Number 1

Estuarine mean flow estimation revisited: Application to the St. Lawrence estuary

by V. G. Koutitonsky[1] and M. I. El-Sabh[2]

ABSTRACT

Mean value estimate errors for estuarine (and oceanic) parameters which exhibit serial correlation and nonstationarity over a finite record length are discussed. It is shown that when trends are nonlinear over the record length, the mean value estimate error does not decrease with time. Instead, it goes through a minimum and then increases again. An optimal averaging time over which the mean estimate error is minimum is presented. The mean circulation in the lower St. Lawrence estuary is described over a record length of 78.5 days in 1979. Standard, bias, and rms errors in mean value estimates are discussed, and an averaging time yielding a minimum error is suggested for the lower St. Lawrence estuary. New measured features of the mean circulation are: a coastal current flowing downstream near the north shore which deflects to the right at the mouth, and a 2 cm/s inflow in the bottom layer at mid-channel location.

1. Introduction

Estuarine processes are usually described in terms of mean value estimates of appropriate parameters sampled over a finite record length. Whenever estimate errors are computed, two assumptions are often made about the sample observations: (1) they are statistically independent, and (2) they are drawn from a stationary random process. As a result, the estimates are considered to be unbiased and constant in time, with a standard error that decreases as the square root of the number of independent observations. However, recent circulation studies in estuaries have revealed that a

1. Institut National de la Recherche Scientifique—Océanologie, 310 Avenue des Ursulines, Rimouski, Quebec, Canada, G5L 3A1.

2. Département d'Océanographie, Université du Québec à Rimouski, 300 Avenue des Ursulines, Rimouski, Quebec, Canada, G5L 3A1.

large portion of the current variability consists of energetic subtidal fluctuations produced by atmospheric, buoyant, and other local and/or nonlocal forcing (Elliot, 1978; Weisberg, 1976a; Wang and Elliot, 1978; Kjerfve *et al.*, 1978; Walters, 1982; El-Sabh *et al.*, 1982). Typical periodicities range from a few days to a few month or more, the latter bound corresponding usually to the record length. Depending on their nature, these fluctuations can have a two fold effect on mean value estimates of estuarine parameters. The first is to introduce serial correlation in the observations. This reduces the number of independent observations since these are now drawn at time intervals too short to be independent. As a consequence, the standard error increases, or conversely a longer record length is required to achieve the same error tolerance in mean value estimates. Serial correlation in stationary processes was addressed by Weisberg (1976b) for estuarine mean flow estimates, and more generally by Davis (1976; 1977) and Chelton (1983) for multivariate oceanographic regression estimates.

The second effect introduced by low frequency fluctuations is that resulting from nonstationarity over the finite record length of the series being averaged. It will be shown that, for the model time series considered, mean value estimates become biased whenever nonlinear trends exist in the series.

The purpose of this note is to complement the study of Weisberg (1976b) by discussing the bias error introduced in estuarine mean value estimates by nonstationarity over the record length. The formalism of mean value estimate errors is reviewed in Section 2, and applied to describe the mean circulation in the lower St. Lawrence estuary in Section 3. Results are discussed in Section 4.

2. Mean value estimate errors

Consider a sample of N scalar observations $u(n\Delta t)$, $n = 0, \ldots, N-1$, drawn at time intervals Δt from a population with mean $\bar{u}$ and variance σ^2. Since the record length $T = N\Delta t$ is finite the time average $\hat{u}$ computed from:

$$\hat{u} \equiv \langle u(n\Delta t)\rangle = \frac{1}{N}\sum_{n=0}^{N-1} u(n\Delta t) \tag{1}$$

is at most an estimate of $\bar{u}$, with a root mean square (rms) error ϵ defined as:

$$\epsilon = (\psi^2 + \beta^2)^{1/2} \tag{2}$$

where ψ is the standard error and β is the bias error (see Bendat and Piersol, 1971 referred to as BP hereafter, p. 171).

Then consider the following scalar time series model:

$$u(n\Delta t) = u'(n\Delta t) + U(n\Delta t) \tag{3}$$

where u' are stationary random fluctuations with zero mean, superimposed on a slowly varying signal U identified here as a trend. The question addressed in this section is:

what is the rms error of the mean estimate $\hat{u}$ of a model series as defined in Eq. (3) over a finite record length T?

Depending on the presence (or absence) of serial correlation in the series and/or nonstationarity over the record length resulting from time variations in U, four cases will now be considered:

a. Independent observations in a finite stationary series. Here the average $\hat{u}$ is an unbiased estimate of the mean value $\overline{u}$, and the rms error ϵ is given by the standard error ψ, for N independent observations:

$$\boxed{\epsilon = \psi = \hat{\sigma}/(N)^{1/2}} \tag{4}$$

where $\hat{\sigma}$ is the standard deviation estimate given by:

$$\hat{\sigma} = \left[\frac{1}{N-1}\sum_{n=0}^{N-1}(u'(n\Delta t))^2\right]^{1/2}. \tag{5}$$

b. Serial correlation in a finite stationary series. Serial correlation can be identified by the two-sided autocorrelation function $R(\tau)$ defined for time lags $\tau = k\Delta\tau$, $k = -M, \ldots, 0, \ldots, M$ as:

$$R(k\Delta\tau) = \frac{1}{\hat{\sigma}^2}\langle u'(n\Delta t)\cdot u'(n\Delta t + k\Delta\tau)\rangle. \tag{6}$$

It can be shown (BP, p. 173) that the average $\hat{u}$ still remains an unbiased estimate of $\overline{u}$, but that the standard error now becomes a function of $R(k\Delta\tau)$ as:

$$\psi = \left\{\frac{\hat{\sigma}^2}{N\Delta t}\sum_{k=-M}^{M}[1 - (|k|\Delta\tau/N\Delta t)]\,R(k\Delta\tau)\cdot\Delta\tau\right\}^{1/2}. \tag{7}$$

For time lags $|\tau| \ll T$, this expression reduces to:

$$\psi = \left\{\frac{\hat{\sigma}^2}{N\Delta t}\cdot \mathrm{T}\right\}^{1/2} \tag{8}$$

where a characteristic time scale T over which observations remain correlated has been defined as (see Lumley and Panofsky, 1964):

$$\mathrm{T} = \sum_{k=0}^{M} R(k\Delta\tau)\cdot\Delta\tau. \tag{9}$$

An effective number of independent observations can then be found as: $N^* = T/\mathrm{T}$, and the rms error in this case is:

$$\boxed{\epsilon = \psi = \hat{\sigma}/(N^*)^{1/2}}\,. \tag{10}$$

Noting that T is the Fourier transform of the variance spectral density function $S(f)$ at frequency $f = 0$, Weisberg (1976b) formulated the standard error as:

$$\psi^2 = \frac{S(0)}{N\Delta t} \tag{11}$$

from which he concludes that in estuaries the mean estimate error ϵ decreases as the averaging time $N\Delta t$ increases. This holds as long as $\beta = 0$ in Eq. (2).

c. Independent observations in a finite nonstationary series. Most geophysical processes exhibit nonstationarity, that is their statistical properties change with time during the period of observation. Since most time series analysis methods are applicable under the assumption of stationarity, it has become standard practice in oceanography to remove nonstationarity in the form of a trend, and proceed with the analysis of the remaining "stationary" series. Note that in this procedure it is implicitly assumed that the series are modelled as in Eq. (3). Conventional methods to remove trends are: (a) by digital high pass filtering the observations, (b) by subtracting linear regressions taken over short segments chosen by some objective technique, (c) by subtracting a low order polynomial fitted over the entire record length.

Even then, the removal of a trend whose time scale of variability is much longer than the averaging time T does not necessarily modify the biased nature of the mean value estimate. Consider for example the model time series specified in Eq. (3). Assuming that $U(n\Delta t)$ is a deterministic process with continuous second derivatives, and that it is statistically independent of $u'(n\Delta t)$, it is shown in the Appendix that the average $\hat{u}(n\Delta t)$ is a biased estimate of $\bar{u}(n\Delta t)$, both considered now as a function of time. It is also shown that the bias error can be approximated as:

$$\beta \simeq \frac{T^2}{24} \langle U_{tt} \rangle \tag{12}$$

where U_{tt} is the second derivative of U with respect to time. In this case, the rms error becomes:

$$\boxed{\epsilon = [\hat{\sigma}^2/N + \langle U_{tt} \rangle^2 \cdot T^4/576]^{1/2}} \tag{13}$$

d. Serial correlation in a finite nonstationary series. The general case, and perhaps the most commonly encountered in estuarine and coastal waters, is one where time series exhibit both serial correlation, and nonstationarity. Replacing Eqs. (10) and (12) in (3), the rms error for this series is found as:

$$\boxed{\epsilon = [\hat{\sigma}^2/N^* + \langle U_{tt} \rangle^2 \cdot T^4/576]^{1/2}} \tag{14}$$

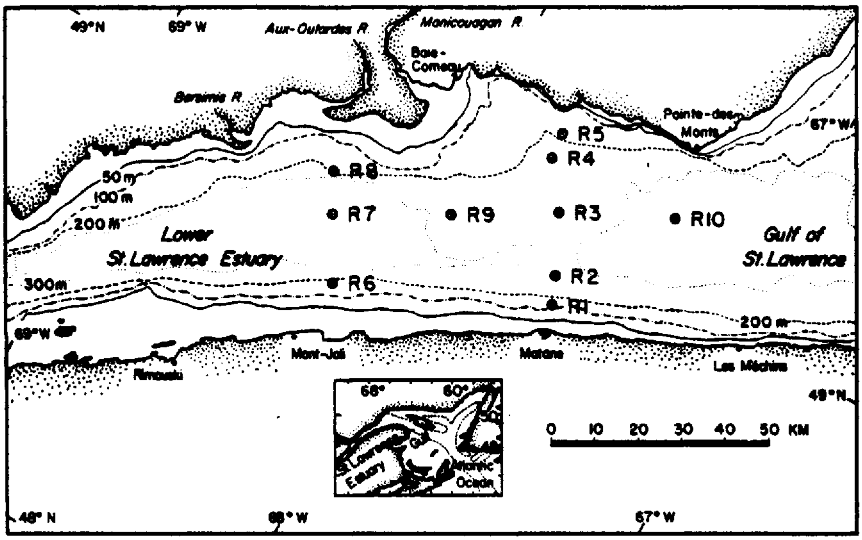

Figure 1. Current meter moorings in the lower St. Lawrence estuary.

Note that in either Eqs. (13) or (14) the rms errors does not necessarily decrease as the averaging time T is increased. In fact, by differentiating either equation with respect to T and setting the result equal to zero, an optimal time T_0 which minimizes ϵ is found to be:

$$T_0 = [144\,\hat{\sigma}^2 \cdot \mathrm{T}/\langle U_{tt}\rangle^2]^{1/5}. \tag{15}$$

For independent observations, T is replaced by $2\Delta t$ above. For stationary observations, or even when a linear trend is present in a series, $U_{tt} = 0$, and T_0 would tend to infinity as expected. Finally, it is suggested that ϵ and T_0 be computed for series from which the main variance-containing semidiurnal and diurnal tides have been filtered out — otherwise the contribution of $\hat{\sigma}^2$ in (13) or (14) is likely to dominate ϵ more than β^2.

3. Mean flow estimates in the lower St. Lawrence estuary

The lower St. Lawrence estuary (LSLE) is an elongated channel with typical length, width, and depth of the order of 200 km, 40 km, and 300 m respectively (Fig. 1). Except for a baroclinic current flowing seaward near the south shore (Neu, 1970), a predominantly southward flow at the mouth (Farquharson, 1966), and a sluggish northward cross-channel motion near Rimouski (Forrester, 1970) features of the steady state circulation in the LSLE remained hypothetical until recently (see El-Sabh, 1979 for a review). In 1979, a large scale circulation study was undertaken in the LSLE. The experiment and the surface residual variability were described by El-Sabh *et al.* (1982). Briefly, Aanderaa current meters recorded water speed,

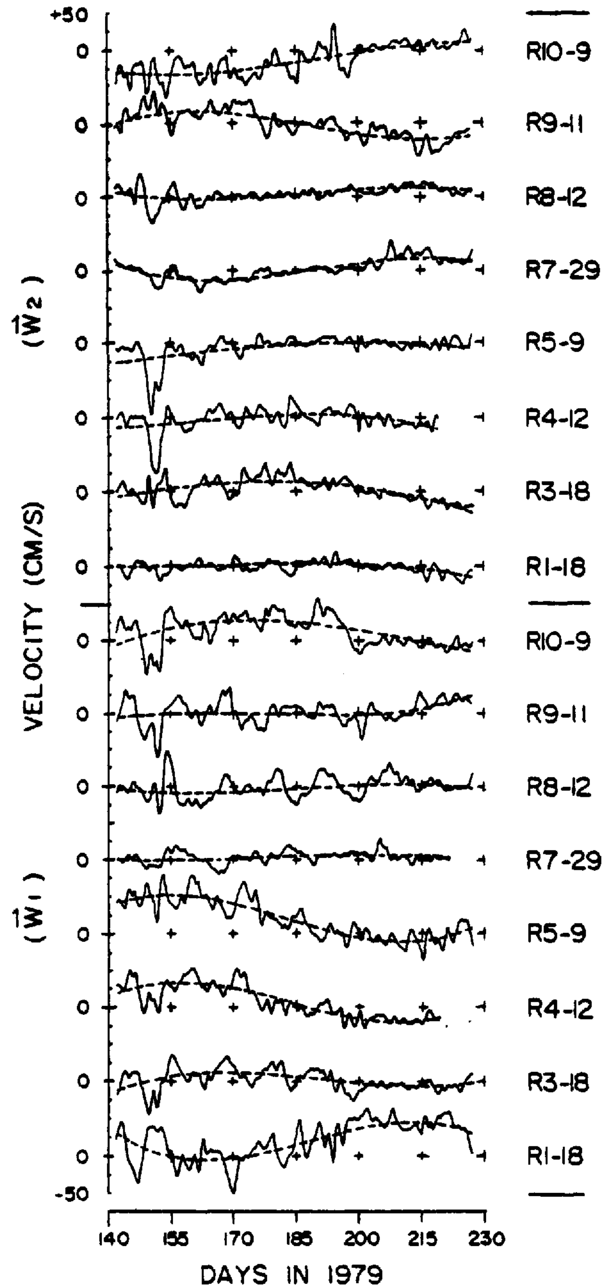

Figure 2. Low pass filtered surface layer velocity component time series, with cubic polynomial trends shown as broken lines.

direction, temperature, and conductivity at mooring sites R1 to R10 (Fig. 1), at various depths and during periods of time lasting from one to six months.

Speed and directions were resolved along orthogonal horizontal axes, X_1 and X_2 oriented positive downstream (66° from true north) and toward the north shore respectively, with corresponding W_1 and W_2 velocity components. All W_i, $i = 1, 2$, time series were then low pass filtered in order to remove deterministic tidal signals and higher frequency noise, using a fast Fourier transform filter (Walters and Heston, 1982) with a cut-off period set at 34 hours. The series were then resampled every 6 hours, and a common record length of 314 points (78.5 days) was retained for this analysis, starting at 0100 hours on May 22 (day 142) 1979. Trends were determined by fitting a univariate curvilinear regression model of degree three to the series. Low passed velocity components W_i, with their trends in broken lines are shown on Figure 2 for eight surface velocity time series.

Visual inspection of all series reveals that velocity components in the LSLE can be modelled as Eq. (3), with stationary 10–15 days periodicity fluctuations u' superimposed on a slowly varying nonlinear trend U. Also the trends do not appear to modulate the fluctuations u', such that they can be considered to be statistically independent of u', as required by Eq. (3).

Standard errors were first computed from Eq. (11). $S(0)$ was taken as the upper 95% confidence interval limit of the variance spectral density estimate of each series (BP, p. 192) at zero frequency, with 14 degrees of freedom. Bias errors were then estimated from Eq. (12), where U_{tt} was obtained from the second derivative of the cubic regression equation. F-value statistics were estimated for each regression, and out of 30 estimates, only 6 were below the value of 10. The rms error for the estimate $\hat{u}$ was finally computed from (14). These errors, as well as the standard deviation of each velocity component series are presented in table (1) as coefficients of variation (percentages of mean value estimates). The averaging times T_0 required for minimum rms errors, and the corresponding errors are also given in Table (1).

4. Discussion

The mean circulation in the LSLE for a period of 78.5 days in 1979 is presented in Figure 3a,b for the surface and deeper layers, respectively. In the upper 30 m, the circulation is characterized by two coastal currents flowing parallel to both shores toward the Gulf with speeds of 20 cm/s. At the mouth, the north shore current seems to veer toward the south shore, producing a cross channel current with speeds over 20 cm/s. As expected from estuarine circulation patterns, most currents in the upper layers flow downstream, with some tendency toward the north shore at mid estuary stations. These mean estimates agree well with the surface circulation sketch proposed by El-Sabh (1979). At mid-depth (M in Fig. 3b) the mean flow decreases and adopts a cyclonic pattern at R2 and R3. Near the bottom, (B in Fig. 3b) the water flows upstream as expected at R3 and R7, with some tendency toward the north shore.

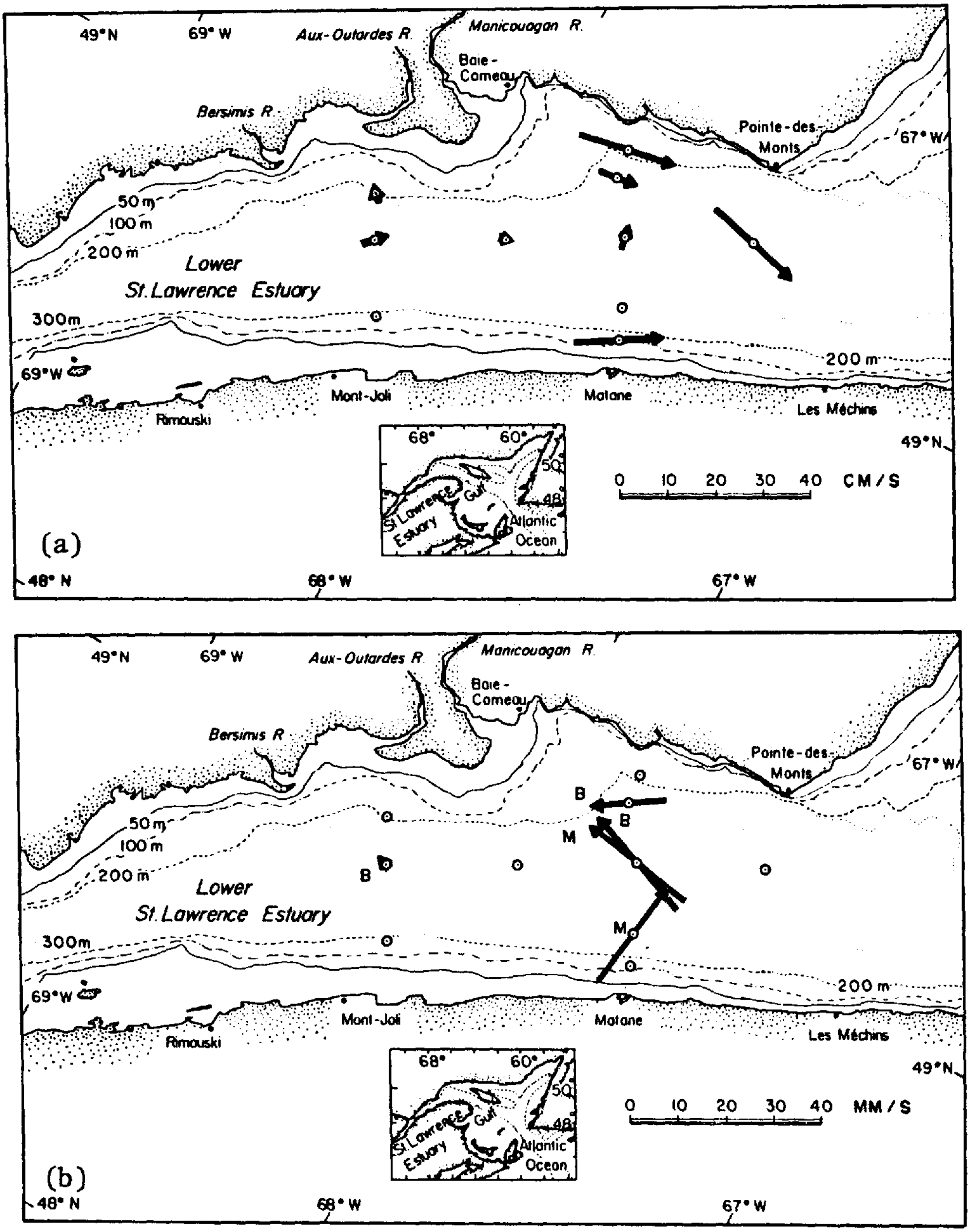

Figure 3. Mean velocity estimates (a) in the upper layer and, (b) in deeper layers where *M* stands for mid-depth and *B* for bottom measurements. Note change of velocity scale in (b).

Table 1. Mean value estimate of velocity components W_1 and W_2, standard errors, bias errors, rms errors, time T_0 for minimum rms error, error at T_0, and standard deviations, in percentages of the mean. All series are 78.5 days long, except for R3B, R2S, R6S, which spread over 71.5, 45, and 30.7 days respectively.

	Mean Estimate (cm/s)		Standard Error (%)		Bias Error (%)		rms error (%)		Time (T_0) for minimum error (days)		rms for $T = T_0$ (%)		Standard Deviation (%)	
Station	W_1	W_2	W_1	W_2	W_1	W_2	W_1	W_2	W_1	W_2	W_1	W_2	W_1	W_2
R1S	18.5	1.0	10	65	46	279	47	287	32	33	17	112	140	733
R3S	1.3	5.0	95	31	396	192	407	194	34	29	163	57	1122	284
R4S	7.9	−2.6	20	62	15	248	25	255	67	34	24	105	277	598
R5S	20.1	−6.4	4	28	6	84	7	88	54	38	6	45	133	253
R7S	5.5	1.9	25	94	53	278	59	293	44	38	37	150	212	1006
R8S	−1.1	3.5	95	25	86	80	128	84	62	38	119	41	1422	301
R9S	1.3	0.2	148	574	240	2684	282	2744	49	32	209	1004	1292	8179
R10S	16.2	−14.6	16	6	75	20	77	21	32	36	27	10	130	135
R2M	1.4	1.9	45	15	173	14	179	21	35	60	76	19	862	221
R3M	−2.0	1.6	25	18	61	181	66	182	41	24	38	37	209	276
R4B	−1.5	−0.1	10	227	125	270	126	353	21	55	21	303	177	2191
R7B	−0.2	0.3	67	68	90	39	112	79	53	74	91	78	895	663
R3B	−1.6	1.9	12	49	3	5	13	49	92	139	12	39	140	550
R2S	1.9	2.7	134	33	316	6	343	34	24	70	204	30	876	379
R6S	22.1	0.2	11	480	19	1366	22	1447	18	15	15	759	67	4184

However, the error analysis presented in Table 1 indicates that only the coastal currents (R1S and R5S), the cross channel flow at the mouth (R10S), and the upstream flow near the bottom (R3B) have rms errors less than 100%. Although standard errors remained below 100% in most cases, high bias errors inflated the rms mean estimate errors well over 100% for currents at other stations. As discussed in Section 2, these errors are attributed to the presence of nonlinear trends in the series. Results of the analysis also show that the rms errors will decrease if averages are computed over shorter time intervals. Most probable averaging times T_0 in Table 1 varied from 35 to 45 days, although the scatter was great. Finally, all standard deviations exceeded 100% of the mean estimate over the 78.5 days, which is indicative of high flow variability in the LSLE.

This study suggests that estimating mean values from time series of estuarine parameters is of little use when trends in the series become time dependent, since bias error will grow as $(T)^4$ while standard errors will decrease as $(T)^{-1}$. Also, a parabolic trend is expected to yield largest bias errors since the mean curvature U_{tt} is maximized. If averages are required under such conditions, the series should be segmented, and averages estimated over segments with linear trends at most. One objective time scale for segmenting time series can be T_0, the time scale over which the rms error goes through a minimum.

It must be stressed however that trends $U(t)$ were considered in Section 2 as being deterministic, which is not necessarily appropriate for estuaries and coastal waters. Trends in general are insidious, and since their probability distribution is not known, statistical inference cannot be drawn. Perhaps one should endeavor to understand the origin of the trend, e.g. buoyancy forcing, etc, and then choose an averaging time accordingly to answer some specific question.

Acknowledgments. We wish to thank Dr. R.E. Wilson from the Marine Sciences Research Center, State University of New York at Stony Brook for his guidance. Thanks are also expressed to Dr. J.P. Chanut from the Université du Québec à Rimouski and two anonymous reviewers for their helpful comments, and to Ms. Marie Cogné for typing the note. This study was supported in part by contract ISD-78-00066 to V.G. Koutitonsky from the Ministry of Supply and Services Canada, and in part by grants from the Fonds FCAC, Québec and the Natural Sciences and Engineering Research Council, Ottawa, to M.I. El-Sabh.

APPENDIX

Bias error in mean value estimates of nonstationary processes

Most of the following material was synthesized from sections in Bendat and Piersol (1971).

Consider a nonstationary process described as:

$$u(t) = u'(t) + U(t) \tag{A1}$$

with variables as in (2c). The mean value of the process at time t is:

$$\begin{aligned}\overline{u}(t) &= E[u(t)] = E[u'(t) + U(t)] \\ &= E[u'(t)] + E[U(t)] = 0 + U(t) \\ &= U(t)\end{aligned} \tag{A2}$$

where $E[\quad]$ is the expected value operator defined for some continuous function $x(k)$ with a probability density function $p(x)$ as:

$$E[x(k)] = \overline{x} = \int_{-\infty}^{\infty} x\, p(x)\, dx. \tag{A3}$$

The mean value estimate $\hat{u}(t)$ over some record length T much shorter than the time variation of $U(t)$ is then found at time t as:

$$\hat{u}(t) = \frac{1}{T}\int_{t-T/2}^{t+T/2} u(t)\, dt = \frac{1}{T}\int_{t-T/2}^{t+T/2} [u'(t) + U(t)]\, dt$$

and

$$E[\hat{u}(t)] = E\left\{\frac{1}{T}\int_{t-T/2}^{t+T/2} [u'(t) + U(t)]\,dt\right\}$$
$$= \frac{1}{T}\int_{t-T/2}^{t+T/2} \left\{E[u'(t)] + E[U(t)]\right\} dt$$
$$= \frac{1}{T}\int_{t-T/2}^{t+T/2} U(t)\,dt. \tag{A4}$$

Since (A4) is not necessarily equal to (A2), the average $\hat{u}(t)$ is a biased estimate of the mean $\overline{u}(t)$. The bias error β can be derived as:

$$\beta(t) = E[\hat{u}(t) - \overline{u}(t)] = E[\hat{u}(t)] - E[\overline{u}(t)]$$
$$= \frac{1}{T}\int_{t-T/2}^{t+T/2} U(\xi)\,d\xi - U(t). \tag{A5}$$

Expanding $U(\xi)$ in a Taylor series about $\xi = t$, and keeping the first three terms,

$$U(\xi) = U(t) + (\xi - t)\,U_t(t) + \frac{(\xi - t)^2}{2}\,U_{tt}(t) \tag{A6}$$

Replacing (A6) in (A5) and integrating:

$$\beta(t) = \frac{1}{T}\left[U(t)T + U_{tt}(t)\frac{T^3}{24}\right] - U(t)$$
$$= \frac{T^2}{24}\,U_{tt}(t). \tag{A7}$$

REFERENCES

Bendat, J. S. and A. G. Piersol. 1971. Random Data: Analysis and Measurement Procedures, John Wiley & Sons Canada Ltd, Rexdale, Ontario, 407.

Chelton, D. 1983. Effects of sampling errors in statistical estimation. Deep-Sea Res., *30*, 1083–1103.

Davis, R. E. 1976. Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. J. Phys. Oceanogr., *6*, 249–266.

—— 1977. Techniques for statistical analysis and prediction of geophysical fluid systems. Geophys. Astrophys. Fluid Dynamics, *8*, 245–277.

Elliot, A. J. 1978. Observations of the meteorologically induced circulation in the Potomac estuary. Estuar. Coast. Mar. Sci., *6*, 285–299.

El-Sabh, M. I. 1979. The lower St. Lawrence estuary as a physical oceanographic system. Le Naturaliste Canadien, *106*, 53–73.

El-Sabh, M. I., H.-J. Lie and V. G. Koutitonsky. 1982. Variability of the near-surface residual current in the lower St. Lawrence estuary. J. Geophys. Res., *87*, 9589–9600.

Farquharson, W. I. 1966. St. Lawrence estuary current surveys. Unpublished manuscript, BIO-66-6, Bedford Institute of Oceanography, Dartmouth, N.S., 84.

Forrester, W. 1970. Geostrophic approximation in the St. Lawrence estuary. Tellus, *22,* 53–65.
Kjerfve, B., J. E. Greer and R. L. Crout. 1978. Low frequency response of estuarine sea level to non-local forcing, *in* Estuarine Interactions, M. L. Wiley, ed., Academic Press, N.Y., 603.
Lumley, J. L. and H. A. Panofsky. 1964. The Structure of Atmospheric Turbulence, Interscience, N.Y., 239.
Neu, H. A. J. 1970. A study on mixing and circulation in the St. Lawrence estuary up to 1964. AOL Rep. 1970–9, Bedford Institute of Oceanography, Dartmouth, N.S., 31.
Walters, R. A. 1982. Low frequency variations in sea level and currents in south San Francisco Bay. J. Phys. Oceanogr., *12,* 658–668.
Walters, R. A. and C. Heston. 1982. Removing tidal-period variations from time series data using low-pass digital filters. J. Phys. Oceanogr., *12,* 112–115.
Wang, D.-P. and A. J. Elliot. 1978. Non-tidal variability in the Chesapeake Bay and Potomac river: Evidence for non-local forcing. J. Phys. Oceanogr., *8,* 225–232.
Weisberg, R. H. 1976a. The non-tidal flow in the Providence River of Narrangansett Bay: a stochastic approach to estuarine circulation. J. Phys. Oceanogr., *6,* 721–734.
—— 1976b. A note on estuarine mean flow estimation. J. Mar. Res., *34,* 387–394.

Received: *11 July, 1984;* revised: *10 October, 1984.*