IV. COVARIANCE FUNCTION ESTIMATORS

The two most important statistical estimators of a random variable **u** are its

mean value $\equiv E\{u\} = \bar{u}$

and

variance $\equiv E\{(u - \bar{u})^2\} = \sigma^2$

This chapter deals with the statistical estimators relevant to the **covariance** of time-domain random variables.

A. Covariance

Consider a simple joint process involving a single-variable ensemble of different wave-induced eastward current measurements shown in Figure IV.1.

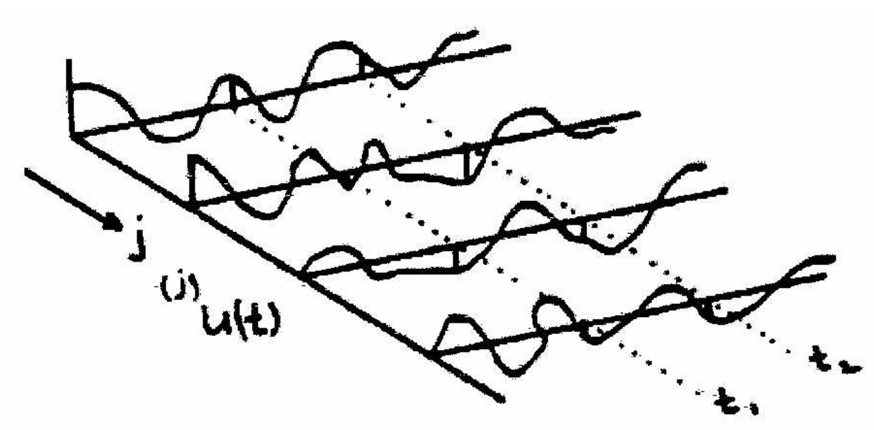

Figure IV.1 A ensemble of wave-induced eastward current measurement realizations. Time horizons at t_1 and t_2 are indicated.

Consider the following *two sets of random numbers*:

$^{(j)}u(t_1)$; one set of random numbers over realizations j

$^{(j)}u(t_2)$; another set of random numbers over realizations j - but at a different time

We define the **joint event $^{(j)}u(t_1)\ ^{(j)}u(t_2)$** (i.e., the same variable at different times). Recalling our definition of probability

$$\Pr\{u(t_1) \subset U_1, U_1 + dU_1\} = f_u(U_1, t_1)\, dU_1 \quad , \tag{IV.1}$$

the **probability of the joint event** is

$$\{u(t_1) \subset U_1, U_1 + dU_1;\ u(t_2) \subset U_2, U_2 + dU_2\} = f_{uu}(U_1, U_2; t_1, t_2)\, dU_1\, dU_2, \tag{IV.2}$$

where $u_1 = u(t_1)$; $u_2 = u(t_2)$.

Recall that the definition of the **expectancy of $u(t_1)$** [or the mean value of $u(t_1)$] is an **ensemble average** of $^{(j)}u(t_1)$ over all j's as given by

$$E\{u(t_1)\} = \int_{-\infty}^{\infty} u_1 [f_u(u_1, t)]\, du_1 = \langle u(t_1) \rangle \quad , \tag{IV.3}$$

where the expectation process can be *expressed symbolically in two ways*, namely E{..} or <..>.

<u>**Joint Covariance**</u> – Now consider the **expectancy of the joint event** - $u(t_1)\, u(t_2)$, which assuming zero mean values $\langle u(t_1)\rangle = \langle u(t_2)\rangle = 0$, is

$$E\{u(t_1)\, u(t_2)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u_1 u_2 f_{uu}(u_1, u_2; t_1, t_2)\, du_1\, du_2 \tag{IV.4a}$$

$$= \langle u(t_1)\, u(t_2) \rangle \tag{IV.4b}$$

Assume that we are dealing with a **stationary process** so that the joint probability density function $f_{uu}(u_1,u_2;\ t_1,\ t_2)$ is independent of absolute time t_1. Thus it can be written in terms of only time differences $t_2 - t_1 = \tau$ (or $t_2 = t_1 + \tau$) according to

$$f_{uu}(u_1, u_2; t_1, t_2) = f_{uu}(u_1, u_2; t_2 - t_1) = f_{uu}(u_1, u_2; \tau). \tag{IV.5}$$

Thus for stationary processes the expectancy of the joint event or **covariance function** in (IV.4) can be written

$$E\{u(t)\, u(t+\tau)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u_1 u_2 [f_{uu}(u_1, u_2; \tau)]\, du_1\, du_2 \,. \tag{IV.6}$$

B. Auto-covariance Functions

In the special case where only a **single random variable u** of a stationary process is being considered, the covariance function in Eq (IV.6) is called an **auto-covariance function**; given by

$$R_{uu}(\tau) = \int_{-\infty}^{\infty}\int u_1(t)\, u_2(t+\tau)\, f_{uu}(u_1, u_2; \tau)\, du_1\, du_2 \,, \tag{IV.7a}$$

or more simply as

$$R_{uu}(\tau) = < u(t)\, u(t+\tau) > \tag{IV.7b}$$

The auto-covariance function $\mathbf{R_{uu}(\tau)}$ is an **even function.** This can be seen by letting $t = t_1 - \tau$ in (IV.7b) which, since the order of the multiplication inside <...> makes no difference, becomes

$$R_{uu}(\tau) = < u(t)\, u(t+\tau) > = < u(t_1 - \tau)\, u(t_1) > = < u(t_1)\, u(t_1 - \tau) > = R_{uu}(-\tau) \ . \tag{IV.8a}$$

or

$$R_{uu}(\tau) = R_{uu}(-\tau) \quad , \tag{IV.8b}$$

which is symmetric and thus an even function.

The normalized form of the auto-covariance function, called the **autocorrelation function,** is

$$\rho_u(\tau) = \frac{R_{uu}(\tau)}{R_{uu}(0)} \quad , \tag{IV.9}$$

where $R_{uu}(0) = R_{uu}(\tau = 0)$.

****

Auto-Covariance Function Example

****

Given the random function

$$^{(j)}u(t) = \sin \omega_o t + {}^{(j)}a \quad , \tag{IV.10}$$

where $^{(j)}\mathbf{a}$ is a *uniformly-distributed* random variable between -π and +π with the PDF shown in Figure IV.2 .

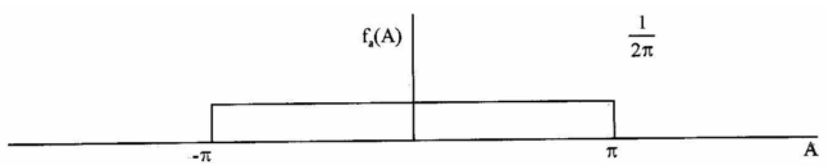

Figure IV.2. The uniform probability density function for phases {a}.

The probability of finding a particular phase value is given by

$$\Pr\{a \subset A, A + dA\} = \frac{dA}{2\pi} = f_a(A)\, dA \,. \tag{IV.11}$$

The auto-covariance function, which involves ensemble-averaging over the j realizations of **a,** is

$$R_{uu} = < u(t)\, u(t+\tau) > = \overbrace{< \sin(\omega_o t + a) \sin[\omega_o (t+\tau) + a] >}^{L} \tag{IV.12}$$

Expanding the right-hand-side of (IV.12) via trigonometric identity yields

$$L = < \sin(\omega_o t + a)[\sin(\omega_o t + a) \cos \omega_o \tau + \cos(\omega_o t + a) \sin \omega_o \tau] > \;. \tag{IV.13}$$

Since τ doesn't vary from realization to realization, the $\sin \omega_o\tau$ and $\cos \omega_o\tau$ terms can be taken outside the ensemble averaging brackets yielding

$$L = [< \sin^2(\omega_o t + a) > \cos \omega_o \tau + < \sin(\omega_o t + a) \cos(\omega_o t + a) > \sin \omega_o \tau] \;. \tag{IV.14}$$

The auto-covariance in (IV.12) is then an ensemble-averaging process, involving an integration over A according to

$$R_{uu}(\tau) = E\{u(t)\,u(t+\tau)\} = \int_{-\pi}^{\pi} [L]\, f_a(A)\, dA = \frac{1}{2\pi}\left(\frac{2\pi}{2}\cos\omega_o\tau\right) = \frac{1}{2}\cos\omega_0\tau. \quad , \qquad \text{(IV.15)}$$

while recalling such trigonometric identities $\sin^2 x = ½\,(1-\cos 2x)$ and $\sin x \cos x = ½ \sin 2x$.
Thus the auto-covariance function of a random trigonometric function is itself a trigonometric function as shown in Figure IV.3.

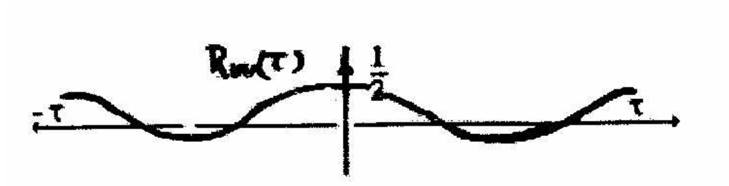

Figure IV.3. Auto-covariance function of a random sinusoid function

****

<u>NOTE:</u> Alternatively, the auto-covariance function given by Eq (IV.15) could have been obtained by taking a <u>time-average</u> over some multiple of the period of the sinusoid; which is true for any <u>stationary</u> process! ……that one can make an ensemble from a single experiment because the process does not "remember" absolute time. This result is particularly important in oceanography because *we rarely have an ensemble of measurements!* Thus we usually must make the assumption of an ergodic process and do time-avergaes.

Auto-covariance Function Estimators

The **true auto-covariance function** of a variable that is part of an **ergodic** and **stationary** process, with an zero mean ($\mu_o = 0$), is

$$R_{uu}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} u(t) \bullet u(t+\tau)\, dt \cdot \qquad \text{(IV.16)}$$

Note that because $T\to\infty$, τ can also $\to\infty$.

For a finite-length series, the **sample auto-covariance function** $R'_{uu}(\tau)$, which approximates $R_{uu}(\tau)$ in the interval $|\tau| \le \tau_{max} \le T$ with τ_{max} the maximum lag, is given by

$$R'_{uu}(\tau) = \frac{1}{T - |\tau|} \int_{\tau}^{T} u(t)\ u(t+\tau)\, dt \qquad \text{(IV.17)}$$

where it is assumed that $R'_{uu}(\tau) \approx 0 \text{.......for} |\tau| > \tau_{max}$. This computation is illustrated in Figure IV.4.

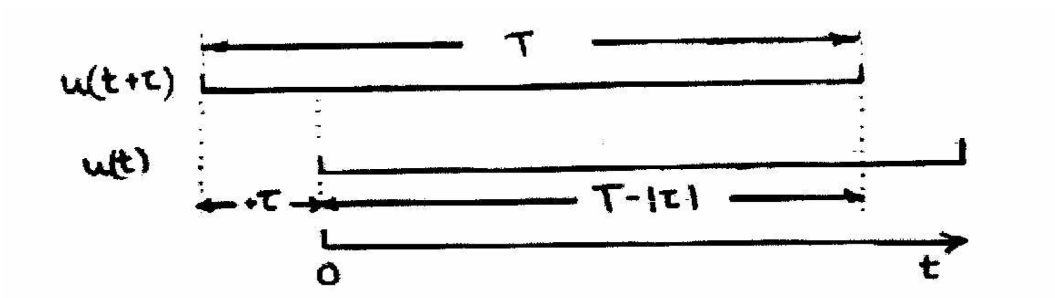

Figure IV.4 The timing of the u(t) auto-covariance calculation.

For a **discrete series u_t**, the **sample auto-covariance function** with zero mean (i.e., $\overline{u_t} = 0$), which approximates $R'_{uu}(\tau)$, is

$$\hat{R}_{uu}(m\Delta t) = \frac{1}{[(N-m)\Delta t)]} \sum_{i=1}^{N-m} u_i\, u_{i+m}\, \Delta t, \qquad \text{(IV.18)}$$

in which $m\Delta t = \tau$ is the **lag,** m the **lag number** in the range $0 \le m \le M$, and $M\Delta t = \boldsymbol{\tau_{max}}$ is the maximum lag.

Eq (IV.18) reduces to

$$\hat{R}_{uu}(m\Delta\tau) = \frac{1}{N-m}\sum_{i=1}^{N-m} u_i\, u_{i+m} \ , \qquad \text{(IV.19)}$$

in which there are only N-m non-zero products in the finite sum.

The sample auto-covariance in (IV.19), with a *non-zero mean* $(\overline{u_t} \neq 0)$, can be written as

$$\hat{R}_{uu}(m\Delta\tau) = \frac{1}{(N-m)}\sum_{i=1}^{N-m} (u_i - \overline{u_t})\,(u_{i+m} - \overline{u_t}) \quad . \qquad \text{(IV.20)}$$

This step reveals a very important property of the auto-covariance function at zero lag (mΔt = τ =0), namely that it is equal to the series total variance $\hat{\sigma}_u^2$ according to

$$\hat{R}_{uu}(0) = \frac{1}{N}\sum_{i=1}^{N} (u_i - \overline{u_t})^2 \quad . \qquad \text{(IV.21)}$$

The normalized auto-covariance function or **auto-correlation function** corresponding to (IV.20) is given by

$$\hat{\rho}_{uu}(m\Delta\tau) = \frac{1}{\hat{R}_{uu}(0)(N-m)}\sum_{i=1}^{N-m} (u_i - \overline{u_t})(u_{i+m} - \overline{u_t}) \qquad \text{(IV.22)}$$

It is convenient to plot **auto-correlograms** because $|\hat{\rho}_{uu}| \leq 1$.

B. Cross-Covariance Functions

General Considerations

Consider a simple joint process, two-signal ensemble of the different current component measurements of a wave field shown schematically in Figure IV.5.

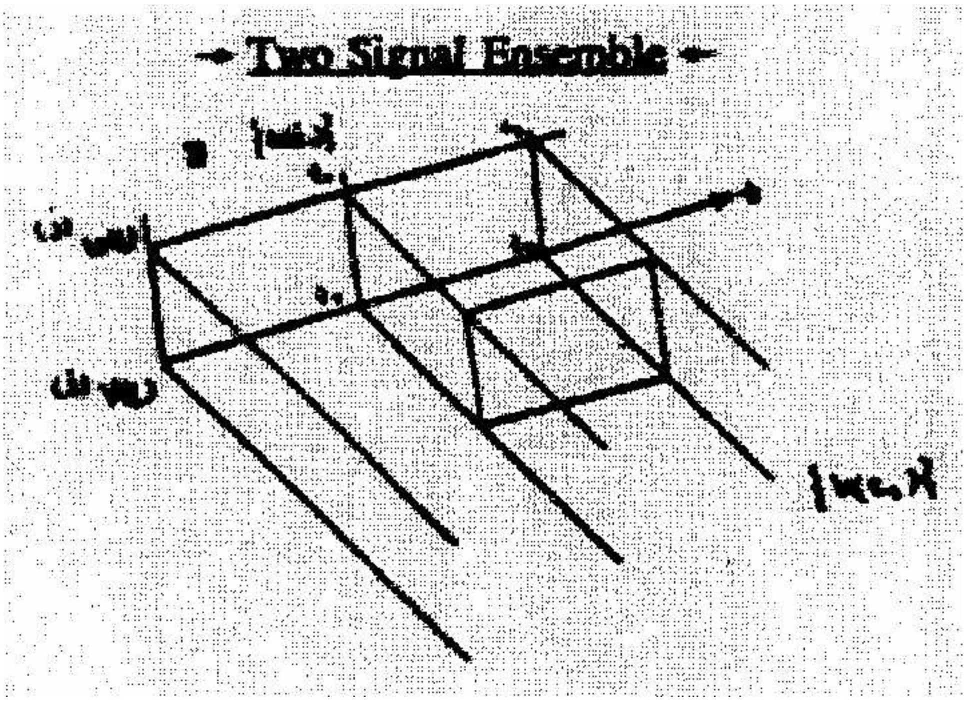

Figure IV.5. An ensemble of eastward u and northward v current measurement realizations.
Time horizons at t_1 and t_2 are indicated.

Consider the following *two sets of random numbers*:

$^{(j)}u(t_1)$; one set of random numbers over realizations j

$^{(j)}v(t_2)$; another set of random numbers over realizations j - but at a different time

We define the **joint event** $^{(j)}\mathbf{u(t_1)}\ ^{(j)}\mathbf{v(t_2)}$ (i.e., different variables at different times). Recalling our definition of joint probability

$$\Pr\ \{u(t_1) \subset U_1, U_1 + dU_1\} = f_u(U_1, t_1)\ dU_1 \qquad , \qquad \text{(IV.23)}$$

the **probability of this joint event** is

$$\Pr\{(u(t_1) \subset U_1, dU_1; v(t_2) \subset V_1, dV_1\} = f_{uv}\,(U_1, V_1; t_1, t_2)\, dU_1 dV_1 \qquad \text{(IV.24)}$$

where $u_1 = u(t_1)$; $v_1 = v(t_2)$.

<u>Joint Covariance</u> – Now consider the **expectancy of the joint event - $u(t_1)\,v(t_2)$.** Assuming for this case that the mean values **$\langle u(t_1)\rangle = \langle v(t_2)\rangle = 0$,** it is

$$E\{u(t_1)\,v(t_2)\} = R_{uv}(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U_1 V_1 f_{uv}(U_1, V_1; t_1, t_2)\, dV_1\, dU_1 \qquad \text{(IV.25a)}$$

$$\langle u(t_1)\,v(t_2)\rangle = R_{uv}(t_1, t_2) \qquad \text{(IV.25b)}$$

If the **joint process is stationary**, then it is only a function of $t_2 - t_1 = \tau$, and the **cross-covariance function** is defined.

$$\langle u(t_1)\,v(t_2)\rangle = \langle u(t)\,v(t+\tau)\rangle = R_{uv}(\tau). \qquad \text{(IV.26)}$$

If $t = t_1 - \tau$, then

$$R_{uv}(\tau) = \langle u(t)\,v(t+\tau)\rangle = \langle v(t_1)\,u(t_1-\tau)\rangle = R_{vu}(-\tau). \qquad \text{(IV.27)}$$

Thus the cross-covariance function has the property that

$$R_{vu}(-\tau) = R_{uv}(\tau) \qquad \text{(IV.28)}$$

which is not necessarily symmetric or even.

The corresponding **cross-correlation function** is defined as

$$\rho_{uv}(\tau) = R_{uv}(\tau)/(R_{uu}(0)\,R_{vv}(0))^{1/2}. \qquad \text{(IV.29)}$$

Note that $\rho_{uv}(\tau)$ will be generally <1.

Sample Cross-Covariance and Correlation Estimators

Consider two discrete records from which trends and other non-stationarties are removed

$$\begin{aligned} u_n &= u(t_o + n\,\Delta t) \qquad n = 1, 2, ..., N \\ v_n &= v(t_o + n\,\Delta t) \qquad T = N\,\Delta t \end{aligned} \tag{IV.30}$$

The sample means of these series are:

$$\bar{u} = \frac{1}{N}\sum_{n=1}^{N} u_n \qquad \bar{v} = \frac{1}{N}\sum_{n=1}^{N} v_n \tag{IV.30}$$

The unbiased estimates of the **discrete sample cross-covariance function** is

$$\hat{R}_{uv}(m\,\Delta\tau) = \frac{1}{N-m}\sum_{n=1}^{N-m}(u_n - \bar{u})(v_{n+m} - \bar{v}) \tag{IV.31}$$

The corresponding sample **discrete sample cross-correlation function** is

$$\hat{\rho}_{uv}(m\,\Delta\tau) = \frac{\hat{R}_{uv}(m\Delta\tau)}{[\hat{R}_{uu}(0)\,\hat{R}_{vv}(0)]^{1/2}} \tag{IV.32}$$

These can be calculated straightforwardly using CORREL) .

****

CHAPTER IV Exercises

****
****

MAR 670 - Data Analysis Methods

(due 28 February 2006)

****

Exercise 2. Auto- & Cross- Covariance and Correlation Functions

a) Compute the following auto-correlation functions using **l_correl**:
50 lags (with (Δt = 1 hr)

input series	**------------->**	**output series**
noise1	------------->	noise1.cuu
sine.10	------------->	sin_10.cuu
sine10.n1	------------->	sn10n1.cuu
sine10.n2	------------->	sn10n2.cuu
sine100.n2	------------->	sn1002.cuu
sinec.n2	------------->	sncn2.cuu
sine10.n3	------------->	sn10n3.cuu

b) Plot the autocorrelations between between -1 and +1 on time scale 0 to 50hr.
(Suggested Scales: vertical = 0.4 units per inch; horizontal = 20 hours per inch).

c) Discuss your results being as quantitative as you can.
For example, why the auto-correlations for
noise1 are so large at $\tau = 0$?; so small at $\tau > 0$?;
sine.10, sine10.n1 and sine10.n2 are so different?; so similar?;
sine10.n2 and sine100.n2 are so different?; so similar?;
sine10.n3 and sine10.n2 are so different ?

d) Compute the following cross-correlation functions using **l_correl:**

Inputs u	**& v**	**---->**	**Cross-Correlations**
noise1	noise2	---->	nos_12.cuv
sine.10	sine.100	---->	s10100.cuv
sine10.n1	sine10.n2	---->	s10_12.cuv
sine10.n1	sine100.n2	---->	s10013.cuv
sine10.n1	sinec.n2	---->	s10c12.cuv
sine10.n1	sine10.n3	---->	s1013.cuv

e) Plot cross-correlation functions using the scales in (b), BUT over the range -50hr to +50hr.

f) Discuss the cross-correlation results, given the variances of the different signal and noise components and the definition of an auto-correlation, both qualitatively and quantitatively.