Goldstone Wind Speeds for the SENSMOD Program

R. Levy
DSN Engineering Section

G. Lorden
California Institute of Technology

Sample months of hourly wind speeds, constructed from two years of observed data in combination with adjustments to reflect a longer data base, are generated for the Goldstone Mars antenna site. The adjustments are determined using a 14-year data base that is available for another site and determining the correlation at the two sites for contemporary periods of observation.

I. Introduction

This paper describes the data analysis and estimation procedures used to establish sample annual records of Goldstone hourly average wind speeds (24 × 365 records per year) from a data base assembled from Goldstone Deep Space Communications Complex and Edwards Air Force Base records. These wind speed records are established to provide input for the SENSMOD[1] energy analysis program (Ref. 1), which is being developed as part of the Goldstone Energy System Project.

The objective is to determine 12 probability distributions of wind speeds, one for each calendar month. The developed distribution for any particular month is to be a representation of the historical speed distribution for that month at the Goldstone site. Samples (including stratified random samples with reduced variance) can then be generated from these 12 distributions to yield independent measurements on an hourly basis and combined to yield "sample years" of wind speed data. No requirement was imposed to model the correlations of successive hourly measurements, which would have imposed a substantial additional program. Consequently, with the present simplification, the distributions of wind speeds are not representative of time intervals shorter than one-month duration.

The form of the distribution function of wind speeds for a given month was assumed to be

$$Prob(Speed \leq s) = F(s) = 1 - \exp\left[-(as + bs^2)\right] \quad \text{for } s \geq 0 \tag{1}$$

where a and b are positive parameters characteristic of the month. Thus, for the ith month ($i = 1, \cdots, 12$), two parameters, a_i and b_i, are to be determined. This form of distribution has been found to provide good fits for the available data sets.

The data base consisted of the following three groups of wind speed records:

Set	Location	Anemometer height, ft	Period of record, yr/mo to yr/mo	No. of entries
A	Goldstone, Mars site	150	66/11–68/10	9100
B	Edwards AFB	13	66/11–68/10	17,500
C	Edwards AFB	13	57/1 –70/12	122,700

The Goldstone data consisted of average wind speeds for 5-min periods. The sampling rate was hourly except for December 1966 through August 1967, when the rate was 1 every 3 hours. Considering the number of observations that could have been made at these two sampling rates during the entire period, it has been found that about 25% of the data is missing. Sets B and C (Edwards data) were extracted from United States Weather Bureau archive tapes (by Meteorology Research Corporation at our request) and are complete sets of hourly records with no entries missing. Set B is actually a subset of C and was assembled for comparison with contemporary Goldstone data.

Since the actual data for Goldstone, set A, covers a period of less than two years, it is not necessarily a reliable basis for the desired probability models. As a means of assuring that the models represent long-term Goldstone phenomena, the following steps were taken, incorporating data sets B and C into the analysis:

(1) Using a chi-square criterion for goodness of fit, 12 sets (a_i, b_i) of monthly parameters were chosen based on the A data.

(2) Using data sets A and B, the degree of correlation between monthly wind power (= speed^3) at Goldstone and Edwards was estimated.

(3) Based on the results of step 2, the long-term data set C from Edwards was used together with A and B to obtain "corrections" of the parameter values found in step 1, the corrected values being designed to incorporate the long-term information provided by set C.

II. Computation Methods

A. Procedure 1 — Generation of Sample from the Distribution of Eq. (1)

A simulated monthly sample of wind speeds s, sample size N (1 per hour), can be constructed from a given distribution by drawing

$$u_i = \text{uniform } (0,1)$$

$$i = 1,2, \cdots, N \qquad (28 \times 24 \leq N \leq 31 \times 24)$$

and letting

$$s_i = F^{-1}(u_i)$$

where F^{-1} is the inverse distribution function. Using the inverse function of Eq. (1) will provide

$$s_i = \frac{(a^2 - 4bc)^{1/2} - a}{2b}$$

where

$$c = \ln(u_i)$$

If desired, variance of the samples so constructed can be reduced by dividing the range (0,1) into p strata and drawing N/p uniform numbers for each stratum, then finding the inverse functions of these numbers as before. The variance of the sample can be made to approach zero by letting p approach N. When p equals N, the sample is no longer a random simulation, but instead tends to replicate the generating distribution. Note that when stratified, the samples will occur as p ordered clusters of N/p numbers. This ordering can readily be removed later by random scrambling of the sample.

B. Procedure 2 — Long-Term Distribution Function Parameters (Steps 1–3, Section I)

1. Step 1. Prior to the present requirement, field data from Goldstone and other weather stations had been processed by a moderate-size computer program WINDMAPW. This consisted of about 1000 source card images plus references to several JPL 1108 FORTRAN V library and IMSL library subroutines for analysis and plotting. After reading the field data from tapes or files, one of the operations of the program was to fit Pearson and Weibull distributions to the data. For the current requirement, the Weibull fit was dropped and fitting methods for the function of Eq. (1), as will be described, were substituted. With these changes, this version of the program was renamed WINDMAPW. The program currently resides in a catalogued file (52219RIL) on the 1108A computer.

Fitting parameters to empirical data was performed in two substeps. A preliminary regression analysis step was executed to provide class boundaries and starting points for a final chi-square calculation step.

a. Preliminary regression estimate for parameters a and b of a sample. By taking the logarithm of both sides of Eq. (1) and rearranging, we have

$$(a + bs) = -\frac{\ln[1 - F(s)]}{s}$$

Then let $Y_i = a + bs_i$ be the model equation for regression analysis of Y on s, and let

$$\frac{-\ln[1 - F(s_i)]}{s_i}$$

be the observed values of Y_i, where $F(s_i)$ is the empirical distribution function of S.

The regression is performed in a subroutine FITEXP. The input to this subroutine is

(1) Sample = $S(I), I = 1,NT$
where NT = total number of terms in sample and S has been sorted in ascending order.

(2) $XMIN$, $XMAX$, = lower and upper threshold speeds that bound the speed region of interest. The values used were 8 and 40 km/h (5 and 25 mi/h) based on typical low-speed threshold and high-speed generator capacity level, respectively.

This subroutine performs the following operations:

(1) Find $IBOT$, $ITOP$, indices of S such that

$$S(IBOT) \geq XMIN$$

$$S(ITOP) \leq XMAX$$

(2) Perform regression on $S(I)$ ($IBOT \leq I \leq ITOP$), using the following algorithm:

(a) $FX = I/(NT + 1)$ = empirical distribution

(b) $X = S(I)$

(c) $Y = -\ln(1 - FX)/X$

(d) $SY = \Sigma Y$

$$SX = \Sigma X$$

$$SXY = \Sigma XY$$

$$SXX = \Sigma X^2$$

Then:

$$NET = ITOP - IBOT + 1$$

$$\mathrm{DET} = NET*SXX - SX*SX$$

The resulting initial estimates for a and b are:

$$a = (SXX*SY - SX*SXY)/DET$$

$$b = (NET*SXY - SX*SY)/DET$$

b. Final chi-square calculation estimate for parameters a, b *of a sample.* Divide the speed range between $XMIN$ and $XMAX$ into 8 interior classes and add one class at the low end to include all speeds below $XMIN$ and one class at the high end to include all speeds above $XMAX$.

Let 0_i = observed number of speeds in class "i" (from the empirical sample)

E_i = expected number in class "i"

Let B_i = lower speed class boundary for class "i"; then

$$E_i = NT * [F(B_i + 1) - F(B_i)]$$

where F(.) is found by evaluating Eq. (1) using the current estimates of a and b.

$$\text{Let } FUNCT(a,b) = \Sigma\,(0_i - E_i)^2/E_i \qquad i = 1, 2, \cdots, 10 \tag{2}$$

The estimate of a and b is made by choosing a, b to minimize the function in Eq. (2).

The operations are performed by subroutines FITEXP and CHIPAM.

FITEXP

(1) After regression has been performed to obtain the preliminary estimate of a, b, classes are set up for the calculation of $FUNCT$ (Eq. 2) as follows: Let NCHI = number of interior classes between $XMAX$ and $XMIN$ (typically 8). Using Eq. (1) compute the difference in distribution function between $XMAX$ and $XMIN$. This difference is divided into $NCHI$ parts, each part representing an equal probability of occurrence. Beginning with

$F(XMIN)$ and adding these part differences successively is equivalent to evaluating the current distribution function at the interior class boundaries. Consequently, evaluation of the inverse distribution function (see Eq. 1) furnishes the class boundaries. At this point, these classes have been established so that $E_i = E_j$ for all the i, j interior classes. However, as the estimates of a and b change from minimization of Eq. (2), the equality is not maintained because class boundaries are not recomputed.

(2) The number of terms in each of these classes is computed (0_i)

(3) Calls subroutine CHIPAM.

CHIPAM

CHIPAM calls a JPL library subroutine to perform a conjugate direction search to find a and b that minimize Eq. (2). The search is performed by means of many (50 to 100) evaluations of the function in Eq. (2). At each evaluation, the E_i are computed using the current values of a and b. The search terminates based upon tolerance tests of 1×10^{-3} for a, 1×10^{-4} for b, and 1×10^{-3} on the function. This last tolerance is equivalent to a very small fraction of the actual function (one part in 10,000 to 70,000).

As a comment on the method of obtaining the final chi-square estimate of the parameters, we find no great changes from the preliminary parameters or function values obtained by regression.

2. **Step 2.** The correlation between data sets A and B was analyzed in terms of wind power P_f, defined as a summation of $\min[S,25]^3$ $(=\min[S^3,25^3])$, where S = wind speed in miles per hour. The computational equation is included in Table 1, and the computed values of P_f for each monthly record in sets A and B are given in columns 1 and 4 of Table 1 (Goldstone) and column 5 of Table 2 (Edwards). It is assumed that the true correlation ρ between total power at Goldstone and Edwards in contemporary months is chronologically invariant. It is also assumed that the pair (Goldstone monthly power, Edwards monthly power) has a specific distribution for each calendar month, so that for any specified month the result for any given year is an independent sample from this distribution. It follows from these assumptions that ρ is also the (true) correlation of the pairs of differences, which are defined as follows:

$$(G_{1i} - G_{2i}, E_{1i} - E_{2i}), i = 1, \cdots, 12$$

where G_{1i} and G_{2i} represent Goldstone power in the ith month of year 1 and year 2, respectively, and E_{1i} and E_{2i} are defined correspondingly for Edwards. Because of missing September and October data at Goldstone, Tables 1 and 2 provide only 10 values of i where data from two years are available. These were used to estimate ρ by computation of the sample correlation coefficient of these 10 pairs, yielding the estimate:

$$\hat{\rho} = 0.486$$

This correlation coefficient appears to be sufficiently large to confirm that the approach outlined is reasonable; e.g., there is a statistical connection between the random deviations from monthly norms at the two locations, Goldstone and Edwards.

For the analysis that will be described in step 3 below, it is necessary also to estimate the ratio of the standard deviations of G_i and E_i, which are the monthly powers for the ith month. This ratio is assumed to be independent of i and was estimated by the ratio of the two sample standard deviations that are determined for the 10 values of $G_{1i} - G_{2i}$ and for the 10 values of $E_{1i} - E_{2i}$. The result was the estimate

$$\hat{R} = \sigma_G / \sigma_E = 953/664 = 1.44$$

3. **Step 3.** To determine "corrected" (a_i, b_i) pairs, it is necessary first to utilize the correlation information in a rational way. The approach adopted was the following:

(1) Estimate the long-term average of the Goldstone total power for the ith month, for $i = 1, \cdots, 12$.

(2) Use the results of (1) to correct the values of a_i and b_i obtained by the chi-square fit in step 1.

To obtain the long-term Goldstone power for $i = 1, \cdots, 12$ let

g_i = true mean total power for the ith month (Goldstone)

A_i = average of G_{1i} and G_{2i} (data set A, Table 1, Col. 4)

B_i = average of E_{1i} and E_{2i} (data set B, Table 2, Col. 6)

C_i = average of $E_{1i}, \cdots, E_{14i}$ (data set C, 14 years, Table 2, Col. 7)

(For September and October, A_i and B_i were based on only one months' record, not two.)

The estimate used for g_i was

$$\hat{g}_i = A_i + \hat{\rho}\hat{R}(C_i - B_i) \tag{3}$$

The values of A_i, B_i, C_i and g_i are shown in columns 4, 2, 3, and 5, respectively, in Table 3. By the results of step 2,

$$\hat{\rho}\hat{R} = 0.486 \times 1.44 = 0.7$$

Equation (3) defines the appropriate "least weighted squares" estimator of g_i in terms of the available data, A_i, B_i, and C_i, as shown in the Appendix. This interpretation depends on the correlation ρ and the standard deviation ratio R being *known*. Using estimates instead, as we have done, is natural and reasonable. More exact analysis is not feasible without knowledge of the form of the joint density functions describing wind at Goldstone and Edwards.

Note that the effect of Eq. (3) is to correct the direct estimate A_i of Goldstone wind power by adjusting for the difference between Edwards short-term power in the contemporary period and long-term power. For example, January data at Edwards (Table 2) shows B_i slightly less than C_i (i.e., $1045 < 1101$), indicating that the short-term wind was slightly below the long-term average and suggesting a positive correction (increase) in the estimate of Goldstone wind power for January. Columns 4 and 5 of Table 3 show the estimated power before (2856) and after (2895) the correction was applied.

Having the estimates g_i of Goldstone wind power for the long term, it is necessary now to use these to correct the a_i and b_i previously obtained. To clarify the method, it is helpful to rewrite Eq. (1) in the form

$$F(s) = 1\text{-}exp\left\{-d\left[\frac{s}{c} + \left(\frac{s}{c}\right)^2\right]\right\}$$

where $c = a/b$ and $d = a^2/b$. Written this way, the family of distribution functions is seen to have a "scale factor" parameter c and a "shape" parameter d. For example, doubling all the wind speeds has the effect of doubling c, leaving d unchanged. On the other hand, if c is held fixed while d is changed, the "shape" changes. The shape effect is that, for very large d, the distribution is close to an exponential distribution, whereas for smaller d, it has thinner tails that are more like those of a normal distribution.

The choice of a distribution to represent the ith month at Goldstone was made as follows: Let $d_i = a_i^2/b_i$ (estimated in step 1) and determine c_i so that the average total power agrees with the value g_i estimated in (1). Thus, a model is chosen which predicts total power in accord with the estimate previously derived and, among all sets of parameters satisfying this restriction, the choice is made to yield the same "shape" as was estimated directly in step 1.

Letting a_i^* and b_i^* denote the parameters of the final distribution chosen for the ith month,

$$a_i^* = a_i/c_i$$
$$b_i^* = b_i/c_i^2,$$

which are shown in columns 7 and 8 of Table 3. The values of the c_i's are shown in column 6. Note that most of the c_i's are close to 1, indicating that the corrections made are small and that the short-term Goldstone data are, on the evidence of sets A, B, and C, likely to be representative of the long-term wind parameters at Goldstone.

Figure 1 shows a diagrammatic summary of the logic underlying the generation of the wind speed samples.

Table 1. Fitting parameters and power terms

Month	Year	From observations on individual months			Year	From data pools within corresponding months			Average P_f
		a	b	P_f	Year	a	b	P_f	
Jan	67	0.019700	0.000268	3695	67, 68	0.08927	0.000765	2570	2856
	68	0.085847	0.001345	2017					
Feb	67	0.028782	0.002222	3028	67, 68	0.100611	0.002506	1476	2111
	68	0.109335	0.002781	1194					
Mar	67	0.013132	0.002406	5230	67, 68	0.048574	0.001700	3872	4258
	68	0.047229	0.00133	3285					
Apr	67	0.007951	0.003035	5347	67, 68	0.025546	0.002326	4530	4801
	68	0.030524	0.002258	4255					
May	67	0.021088	0.002853	3966	67, 68	0.034341	0.002600	3722	3782
	68	0.044651	0.003081	3598					
Jun	67	0.026241	0.003175	2986	67, 68	0.023422	0.003705	3248	3158
	68	0.022351	0.004697	3329					
Jul	67	0.000001	0.006994	2097	67, 68	0.003654	0.006917	1907	1928
	68	0.004859	0.006800	1758					
Aug	67	0.058394	0.005100	1475	67, 68	0.043111	0.004707	1981	1875
	68	0.030681	0.005051	2275					
Sep	68	0.070572	0.003518	1681					1681
Oct	68	0.144353	0.000544	1088					1088
Nov	66	0.076671	0.000677	2265	66, 67	0.102385	0.001346	1933	1803
	67	0.132794	0.000640	1341					
Dec	66	0.106911	0.000412	3090	66, 67	0.074673	0.000618	3489	3365
	67	0.068373	0.000811	3631					
				*		*	*		*
				(1)		(2)	(3)		(4)

*Indicates that data in the column above was used in either correlation coefficient computation or in long-term Goldstone projection.

$$P_f = \int min(S^3, 25^3)p(s)ds \simeq \sum min(C_i^3, 25^3)\ p(C_i)\ \Delta C$$

where

C_i = class mark

ΔC = class interval (1 mph)

Table 2. Edwards AFB data

Month	Year	P_f	Power terms			
			Year	Average P_f	Year	Pooled P_f
Jan	67	1021	67, 68	1045	57–70	1101
	68	1070				
Feb	67	728	67, 68	1018		1477
	68	1308				
Mar	67	2679	67, 68	2325		2453
	68	1971				
Apr	67	2831	67, 68	2792		2814
	68	2754				
May	67	2580	67, 68	3124		3377
	68	3665				
Jun	67	2660	67, 68	3064		3179
	68	3467				
Jul	67	1547	67, 68	2058		2061
	68	2568				
Aug	67	958	67, 68	1584		1724
	68	2211				
Sep	68	1532				1464
Oct	68	1225				1211
Nov	66	894	66, 67	750		1088
	67	605				
Dec	66	1624	66, 67	1428		960
	67	1232				
		[a] (5)		[a] (6)		[a] (7)

[a] See Table 1 note.

Table 3. Goldstone long-term projected parameters

	P_f = power term				Speed factor	Long-term parameters	
	Edwards		Goldstone				
①	② 66–68	③ 57–70	④ 66–68	⑤ Projected	⑥ c	⑦ a	⑧ b
Jan	1045	1101	2856	2895	1.0575	0.08442	0.00068
Feb	1018	1477	2111	2432	1.2581	0.07997	0.00158
Mar	2325	2453	4258	4348	1.0776	0.04506	0.00146
Apr	2792	2814	4801	4817	1.0220	0.02500	0.00223
May	3124	3377	3782	3959	1.0303	0.03333	0.00245
Jun	3064	3179	3158	3239	0.9911	0.02363	0.00377
Jul	2058	2061	1928	1930	0.9650	0.00379	0.00743
Aug	1584	1724	1875	1973	1.00045	0.04339	0.00470
Sep	1538*	1464	1681	1633	0.9711	0.01267	0.00373
Oct	1225*	1211	1088	1078	0.9448	0.15278	0.00061
Nov	750	1088	1803	2404	1.1364	0.09010	0.00104
Dec	1428	960	3365	3037	0.9251	0.08072	0.00072
Reference							
	Table 2 (6)	Table 2 (7)	Table 1 (4)	④ + 0.7 [③ − ②]	From computer search	$1/c \times (2)$, Table 1	$(1/c)^2 \times (3)$, Table 1

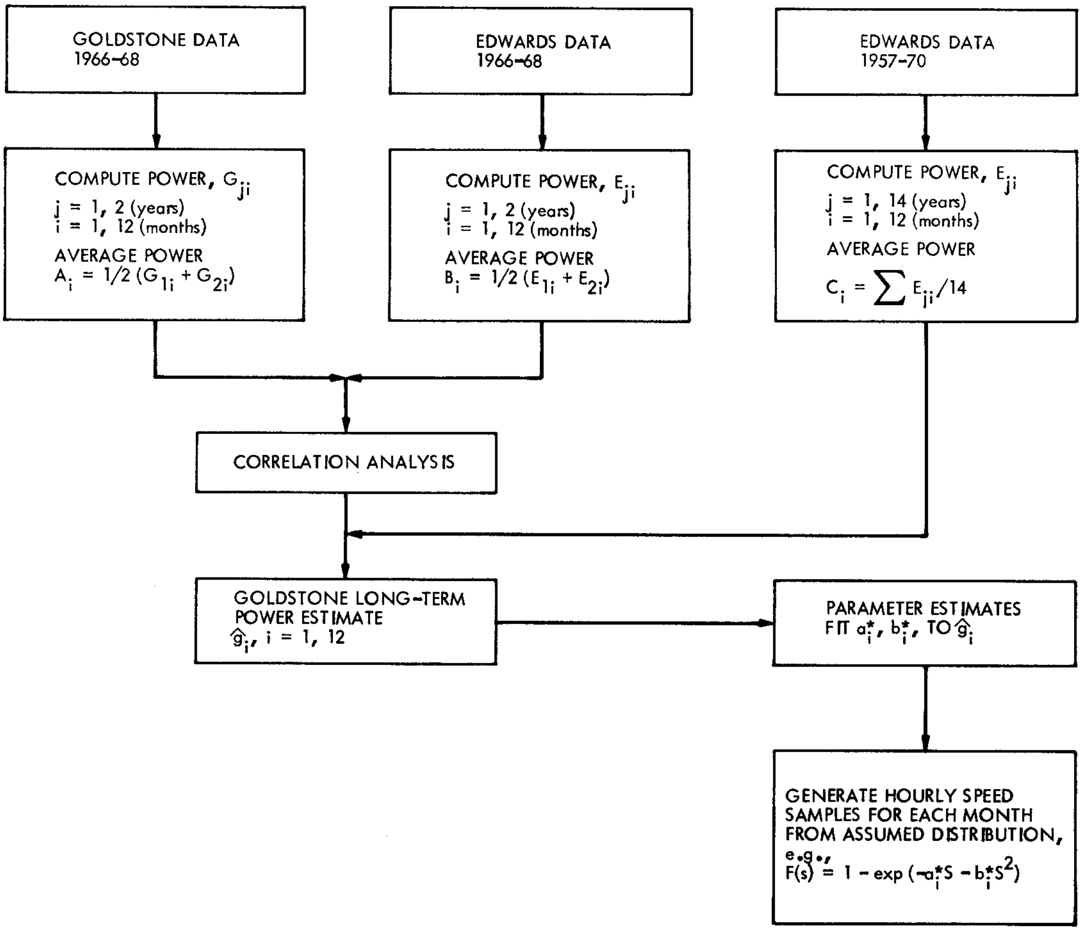

Fig. 1. Goldstone wind speed model for SENSMOD Program

Appendix

Estimation Method

Assume that X and Y are random variables with known correlation coefficient ρ ratio of standard deviations σ_x/σ_y, but with unknown means μ_x and μ_y. We wish to estimate μ_y from data available within two independent samples: the first sample consisting of n paired sets of (X,Y) data, denoted as $(X_1,Y_1), \cdots, (X_n,Y_n)$; the second sample consisting of m observations of a set of X data only, denoted as $X_{n+1}, \cdots, X_{n-m}$. The following sample means can be computed for these data:

$$X_n = \frac{(X_1 + \cdots + X_n)}{n}$$

$$Y_n = \frac{(Y_1 + \cdots + Y_n)}{n}$$

$$X_m = \frac{(X_{n+1} + \cdots X_{n+m})}{m}$$

$$X_{n+m} = \frac{(X_1 + \cdots X_{n+m})}{n+m}$$

The approach here will be to select as the estimator of μ_y the estimate with smallest variance that can be determined from the set of unbiased estimators that can be constructed as linear combinations of the data. In the special case where (X,Y) are normally distributed, this is also the estimator that would be derived by the maximum likelihood method. An alternative derivation would be to use a least squares approach via a transformation to a problem of uncorrelated variables.

As a linear combination of the data, let

$$T = a_1X_1 + \cdots a_nX_n + b_1X_{n+1}$$

$$+ \cdots + b_mX_{n+m} + c_1Y_1$$

$$+ \cdots + c_nY_n$$

be the general form for the estimators to be considered. It will simplify matters to verify first that the variance-minimizing T has all a's equal, all b's equal, and all c's equal.

To see this, consider an unbiased estimator T, not satisfying this condition. By permuting the a's if they are unequal (or the b's or c's if the a's are equal), we can write another estimator $T_2 (\neq T_1)$, which obviously has the same mean and variance as T_1. In particular, T_2 is also unbiased and thus $(T_1+T_2)/2$ is also unbiased.

Now,

$$\frac{Var(T_1+T_2)}{2} = \frac{1}{4} VarT_1 + \frac{1}{4} VarT_2 + \frac{1}{2} Cov(T_1,T_2)$$

$$= \frac{1}{2} VarT_1 + \frac{1}{2\rho} (T_1,T_2)\,(VarT_1)^{1/2}\,(VarT_2)^{1/2}$$

$$= \frac{1}{2}\,[1+\rho(T_1,T_2)]VarT_1$$

which is less than $VarT_1$, unless $\rho(T_1,T_2) = 1$. This last cannot be the case or else there would be a linear relationship between T_1 and T_2, and then by virtue of the equality of their means and variances, T_1 and T_2 would have to be equal. Thus, $(T_1+T_2)/2$ is better than T_1 so that T_1 cannot be the best linear unbiased estimator.

It suffices, therefore, to consider estimators which can be written in the form

$$T = a\overline{X}_n + b\overline{X}_m + c\overline{Y}_n$$

for which the expected value is

$$ET = (a+b)\mu_x + c\mu_y$$

The condition of unbiasedness requires that $a + b = 0$ and $c = 1$. Hence,

$$T = \overline{Y}_n + a(\overline{X}_n - \overline{X}_m)$$

and

$$VarT = Var(\overline{Y}_n + a\overline{X}_n) + Var(a\overline{X}_m)$$

$$= \frac{\sigma_y^2}{n} + \frac{a^2\sigma_x^2}{n} + 2\,Cov(\overline{Y}_n\,\overline{X}_n) + \frac{a^2\sigma_x^2}{m}$$

Since

$$Cov(Y_n,X_n) = \frac{Cov(\overline{Y}_1,\overline{X}_1)}{n} = \frac{\rho\sigma_x\sigma_y}{n}$$

$$VarT = \frac{\sigma_y^2}{n} + a^2\sigma_x^2\left(\frac{\frac{1}{n+1}}{m}\right) + \frac{2a\rho\sigma_x\sigma_y}{n}$$

The minimum is attained at $a = -\rho(\sigma_y/\sigma_x)m/(m+n)$ and equals $\sigma_y^2/n[1-\rho^2 m/(m+n)]$. Using the fact that $\overline{X}_{n+m} = (n\overline{X}_n+m\overline{X}_m)/(m+n)$, the estimator can be written in the form

$$T = \overline{Y}_n-\rho\left(\frac{\sigma_y}{\sigma_x}\right)(\overline{X}_n-\overline{X}_{n+m})$$

Comparing with the estimator $\overline{Y}_n$, which has variance σ_y^2/n, it is seen that the variance is reduced by a factor of $(1-\rho^2)$, when m is much larger than n.

In the application to estimation of monthly wind powers at Goldstone, since the correlation ρ and the standard deviation ratio σ_y/σ_x were unknown, it was necessary to use estimates of these quantities in the formula for T.

Reference

1. Hamilton, C. C., "A Dynamic Model for Analysis of Solar Energy Systems," in *The Deep Space Network Progress Report 42-27*, pp. 41-45, Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1975.