Electric Power Systems Research, 2 (1979) 1 - 14

The Reliability of Distributed Wind Generators

EDWARD KAHN

Energy and Environment Division, Lawrence Berkeley Laboratory, Berkeley, CA 94720 (U.S.A.)

(Received August 16, 1978)

SUMMARY

This paper explores the data and models necessary to evaluate the reliability of wind generators that are geographically distributed in a utility system. A case study is conducted using California data. Results on capacity credit are found to depend upon the size of the region, the reliability of the wind regime, the existing power system generator mix, and the penetration of the wind generators.

1. INTRODUCTION

This paper addresses itself to the reliability of geographically dispersed wind turbine generators. The output of electricity produced from wind machines fluctuates widely at a given site. Yet the value of electricity is a function of its reliability. We typically want electricity available on demand. Pricing schedules have traditionally valued firm power, that is, reliably available power, much more highly than "dump power", which is available intermittently on an "if and when" basis. The conventional wisdom on wind power suggests that it is unrealistic to expect that wind generation will be sufficiently reliable to displace conventional capacity. For example, a recent report [1] expresses the situation as follows:

"Because it is not possible to accurately predict when wind energy will be available, it is usually not practical to assign capacity credit to wind units... . For the majority of utilities, it will be necessary to provide conventional generating facilities as backup... ."

While such conclusions may be valid for the analysis of individual sites, the main thesis of this paper is that geographical dispersal improves aggregate reliability. If the wind is calm at some sites, we can count on it blowing at others. This means that it is reasonable to expect capacity credit, that is, the displacement of conventional capacity, from an array of wind generators spread over a large region.

To examine this thesis we need to outline the methodology required for such assessments and examine data to estimate the magnitude of the dispersal benefit. It is important to emphasize that the analyses presented here are preliminary. The limitations in data and methodology will be discussed at some length. The greatest uncertainties are due to lack of adequate wind resource data. In particular, we need to know how dependent the wind speed at one site is on the wind speed at another. This dependence is called the "statistical correlation". In general, the lower this dependence, the more reliable an array of dispersed wind generators will be. At present we do not know too much about the wind correlations, so our analysis can only present a stylized vision of the subject, limited to a small number of sites. Nonetheless there are enough data to allow the formulation of interesting questions and the statement of preliminary results.

There are also significant problems of method. To address our problem we must marry statistical meteorology with utility methods of power generation reliability assessment. If we do not make some simplifying assumptions, our analysis will drown in a sea of data requirements and endless computation. For conceptual purposes as well, we must find ways of identifying the major variables and labeling our final results.

We shall find that reliability is gained from geographical dispersal, but that there are several limitations to this benefit, two of the most important being the saturation of regional wind diversity and the barrier of large penetrations. The first effect, saturation of diversity, says that as the region examined increases in size, the marginal benefit decreases. From our data we shall find that the reliability aggregated over sites in northern California is only slightly poorer than that of arrays encompassing the entire Pacific Coast region. This is consistent with a similar study of German data which found diminishing marginal benefit over larger and larger wind regimes [2].

The second limitation, the barrier of large penetrations, results from the statistical fact that wind speed correlations never disappear entirely. This fact means that an array of wind generators will act as a single unit in some sense. As more and more such generators are added to a system, *i.e.*, as the penetration increases, the array requires a larger and larger backup. Since the array reliability is never perfect, large penetrations will be marginally less effective in meeting demand. In a sense this diseconomy of scale which has plagued some of the large-unit conventional technologies will also limit wind arrays [3]. From our data we shall find this barrier appearing when wind is roughly 25% of the generator mix.

Because this subject is complex and not widely understood, we shall begin with a quick tour through the methodology. This will simplify the real problems and provide some perspective on the analysis required. After this we shall discuss California data. The limitations of the data will allow us to highlight some additional questions of methodology and to give estimates of the magnitude of dispersal benefits.

2. METHODOLOGY: FROM WIND POWER FREQUENCY TO LOSS-OF-LOAD PROBABILITY AND BACK AGAIN

2.1. *The basic steps*

Wind energy is available intermittently. To represent the variability in some compact form, it is useful to use a cumulative frequency distribution. In Fig. 1 we show such a graph for a single site and a large array. A given point on such a graph associates to a certain level of power the frequency with which power up to that level is available. Thus

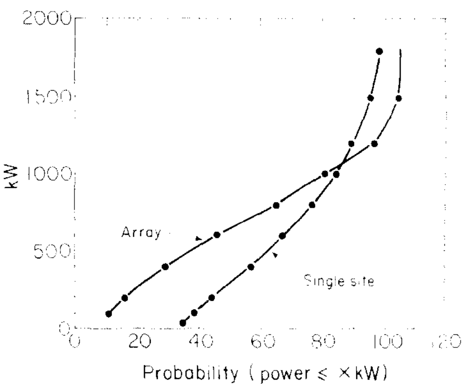

Fig. 1. Power frequency distribution: array *vs.* single site.

the median value for the single site in Fig. 1 is about 300 kW for 2000 kW of generating capacity. For comparison the array value for a frequency of 50% (the median) is about 650 kW per 2000 kW of capacity. Notice that we can only compare array output over the number of sites in the array.

The data represented in Fig. 1 are the first kind of input for a capacity credit analysis. There are many problems associated with getting such data. First, the wind speed data which are presently available are typically collected at heights below that proposed for large wind generators. They must be scaled up for analysis, and that is an uncertain process. For array output curves such as that shown in Fig. 1 we need data on the correlation of wind speed among sites. A brute force calculation can be done if there are good data for all sites. There are also some relatively simple approximations which can be made to array performance. One such technique, developed by C. G. Justus and associates at Georgia Institute of Technology [4], will be employed in our data analysis for California.

Assuming that wind power frequency data for an array are available, the next step in the process is the convolution of this resource with the generators used in the utility system of interest. The calculation of capacity credit is done in the context of a power system generation reliability model. The most common of such models is the "loss-of-load probability" (LOLP) calculation (see ref. 5, for example). This index measures the probability of load being in excess of resources. Generators are represented probabilistically in such models by means of a single unavailability parameter, the forced outage rate.

This is a reasonable portrayal of conventional unit performance; the unit is either on or off. More recently, this representation has been augmented to allow for partial unavailability which can be frequent but usually involves only a small fraction of unit capacity. The curves of Fig. 1 are not easily made compatible with those of standard generator models. To capture the full variability of wind power, many "states of availability" and their probabilities are required. Moreover, the wind power from an array will act as an aggregate. Each wind unit cannot be treated as an independent random variable. In the LOLP calculation each conventional unit is assumed independent. Since the wind generation is correlated, its representation in the LOLP analysis must show this.

After the wind power is given a many-state power availability representation, it can be included with conventional resources to calculate the LOLP. An example calculation is shown graphically in Fig. 2. In this graph we plot LOLP as a function of load for a base case and a case with a large penetration of dispersed wind generators. For utility planning purposes an LOLP objective, the reliability constraint, is imposed. When forecasted load exceeds this specified level, it is a signal for additional capacity requirements. The criterion shown in Fig. 2 is one version of the ubiquitous but unmotivated "one day in ten years" rule of thumb.

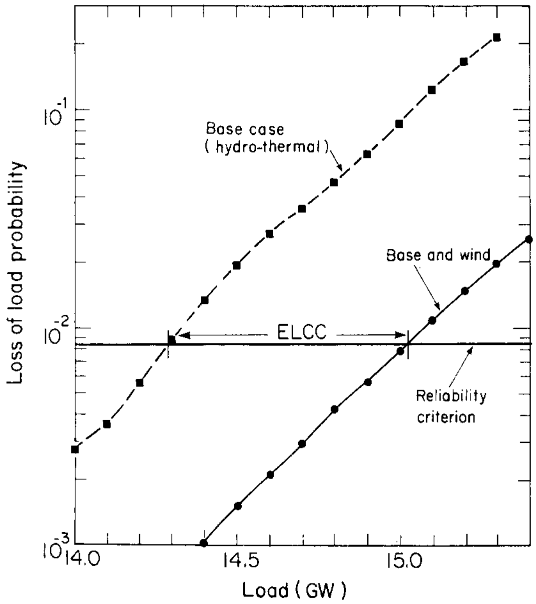

Fig. 2. Definition of effective load carrying capability.

To address the question of capacity credit for wind arrays, we need a common currency in which we can trade off different types of generation. Most useful for such purposes is a notion known in the technical literature as the "effective load carrying capability" (ELCC). This is just the distance between the base case maximum load and the load for the case with wind generation measured at the LOLP equal to the risk criterion. Any unit added to a base system will shift the risk curve to the right. The magnitude of the shift (ELCC) is a function of unit size, forced outage rate, and system characteristics. We can say that one supply alternative displaces another if the ELCC for each is the same. There are, of course, other ways to achieve capacity displacement at constant reliability. These involve backup generation. For example, wind machines can be added in some array configuration displacing other resources with greater ELCC, then gas turbines can be added to the system to restore equal ELCC for the two options. This will involve a greater economic burden than the simple case with just pure dispersal.

2.2. Optimization of various kinds

Starting with the basic tools of a wind power frequency curve and LOLP model one can ask a variety of questions. Clearly there is more than one way to construct a wind array. In the simple calculations to follow we shall assume that all sites are weighted equally. As penetration increases we are just adding more machines at the known locations. This is partly a data constraint in our case but, even with what we know, array performance could be improved by adding more machines at favorable sites. The problem simply is to identify what "favorable" means in this context. If we had negatively correlated sites, they would be favorable. Negative correlation means that when it is windy at site A it is calm at B and *vice versa*. Such perfect complementarity has yet to appear in the data. If it existed, then about half the array capacity should be sited at locations that were negatively correlated with the remaining sites. We shall see in § 3 that there is a tradeoff between high mean wind speed and low positive correlation for array design. Our data provide anecdotal but not conclusive insight into this tradeoff.

Wind regimes that are optimal for power reliability may or may not coincide with administrative districts which currently exist. The dispersal benefit suggests that coordination over large areas is desirable. Small municipal utility systems would probably be unable to capture the diversity. The same is probably true for most typical utility service areas. Large geographically diverse service areas, such as that of the Pacific Gas and Electric Company, encompassing most of northern California, may well be near the optimal size from this point of view. In § 4 we shall examine the question of whether the wind resource diversity as we know it today provides a justification for integrating all the Californian utilities into a regional pool. This question is of independent interest [6, 7].

Our discussion of capacity credit and the previous remarks about the inability of municipal utilities to capture a geographical economy of scale are not conclusive arguments for centralized exploitation of wind energy. An institutionally decentralized implementation of wind machines could still benefit from geographical dispersal if an appropriate mechanism for allocating mutual aid or backup responsibility can be found. Existing power pool agreements are a model of such arrangements. Responsibility and benefit are allocated to members on the basis of investment share and the statistical performance of equipment. Similar tools could be brought to bear in the management of a regional "wind pool". More thorough discussion of such arrangements can be found elsewhere [6]. For our purposes it is important to know that a variety of social and institutional configurations can capture the geographical scale economy of dispersal, if an appropriate statistical foundation can be found to characterize the resource.

There are, of course, many other kinds of optimization involved in integrating wind arrays into an electric energy system besides array design and the optimization of social administration of the resource. Machine design, for example, is guided by principles of optimal resource management. It is important from the systems viewpoint, however, to realize that implementing large wind arrays may require redesign of other elements in the power system. In particular, those regions with substantial hydro resources may find significant opportunities for matching wind energy output fluctuations with hydro dispatch for a smoother total output. Such an operating strategy would probably justify investment in additional turbines for "shaping" of the hydro resource to fit the wind regime. These issues, which are part of the ultimate integration problem, lie beyond our more simple task of illustrating the dispersal benefit with sample data. It is to that task that we now must turn.

3. WIND DATA AND MODELS FOR ARRAY CHARACTERIZATION

3.1. Introduction

Our study of wind diversity in California derives almost entirely from the research of C. G. Justus and associates at Georgia Institute of Technology [4, 8]. The scope and limitations of this work are important to understand before particular results are discussed. Justus has attempted to model the performance of large wind turbine generators such as those proposed by the Department of Energy (DOE). To develop a profile of wind speed frequency distribution at the sites analyzed, the data from 6 m must be scaled up to the 60 m hub-height of the proposed DOE machines. Justus does this by means of some empirical scaling laws associated with the parameters of the Weibull probability distribution. This distribution is particularly useful for wind power analysis, but one must rely on ersatz data for the task at hand. To estimate the error in the scaling laws, a sensitivity analysis should be done using additional upper air data. Our method uses estimates of array reliability that are conservative.

Justus uses data for the Pacific Coast sites listed in Table 1; seasonal variation in mean wind speed is given in Table 2. Among the California sites, five of the six are in Pacific Gas and Electric Company's service area. Although there are some data from more promising sites than those studied by Justus (Pt. Arena and San Gorgonia Pass, see ref. 9), recorded at more appropriate heights, only six months of data are available. This compares with five years for most of the Justus

TABLE 1

West Coast sites

ACV	Arcata, CA	SAC	Sacramento, CA
AST	Astoria, OR	SCK	Stockton, CA
BFL	Bakersfield, CA	SDB	Sandberg, CA
EUG	Eugene, OR	SEA	Seattle/Tacoma, WA
MHS	Mt. Shasta, CA	SFO	San Francisco, CA
NUO	Sunnyvale, CA	SLE	Salem, OR
OTH	North Bend, OR	SMP	Stampede Pass, WA
PDX	Portland, OR	SMX	Santa Maria, CA
RBL	Red Bluff, CA	SXT	Sexton Summit, OR
RDM	Redmon, OR		

Note: California correlation data are available only for BFL, SAC, SFO, SCK, RBL and SDB.

TABLE 2

Mean wind speed (m/s) at 60 m (197 ft) hub-height for pacific sites

	Winter	Spring	Summer	Fall	Annual
AST	7.2	5.6	6.1	6.0	5.8
BFL	4.8	6.0	6.0	5.1	5.5
EUG	6.4	5.5	6.0	5.8	5.5
OTH	7.1	6.3	7.9	6.7	6.3
RBL	6.6	6.2	6.3	6.3	5.8
RDM	6.2	5.2	6.0	5.8	5.2
SAC	5.1	5.7	6.5	4.9	5.7
SCK	5.8	6.6	7.0	5.7	6.3
SDB	9.0	9.5	7.6	8.0	8.5
SEA	6.8	6.1	5.8	6.1	6.2
SFO	5.8	8.1	8.5	6.7	7.3
SMP	7.9	7.4	7.2	7.2	7.5
SMX	5.7	4.8	–	5.7	4.9
SXT	7.9	7.3	6.2	6.8	7.8
Average	6.5	6.5	6.7	6.1	6.2

TABLE 3

Average spatial cross correlation of wind speed: California sites*

	BFL	SAC	SFO	SCK	RBL	SDB
BFL		0.35	0.40	0.37	0.32	0.17
SAC			0.30	0.49	0.29	0.01
SFO				0.44	0.23	0.19
SCK					0.30	0.08
RBL						0.07
SDB						

*See Appendix 1 of ref. 10, July and August values averaged.
Pacific Coast array summer correlation averages 0.25.

sites. Since our primary interest is in geographical correlations, we must rely upon the larger data set. Because there is a large number of data points, we can have reasonable confidence in the correlation coefficients. In Table 3, we summarize data on the summer average correlation among California station pairs. We recall that the correlation coefficient is a measure of linear dependence that varies between +1 (perfect correlation) and −1 (complete inverse correlation). For comparison we also tabulate average monthly spatial cross-correlations for the entire Pacific Coast array. More detailed data are given in Appendix 1 of ref. 10.

As we might expect, Table 3 shows rather high correlation among sites that are close geographically. Thus Sacramento and Stockton show summer correlations of 0.49. Conversely, sites which are relatively remote from one another show lower correlation. San Francisco and Red Bluff have correlations of 0.23. The southern California site, Sandburg, shows low correlations with the northern California stations. In the summer these average about 0.10.

3.2. Weibull approximation model

It would be useful to have a simple model that used the correlation coefficient and a few other parameters to characterize wind arrays. One such model has been developed by Justus. It approximates array performance based on data for a single representative site. For our analysis we rely on data for Sacramento. Individual sites and arrays are characterized by the two parameters of the Weibull distribution, a shape factor k, and a scale factor c. Once these parameters are specified, the probability of power output less than or equal to some value P_j is given by

$$\Pr[P \leqslant P_j] = 1 - \exp[-(V_j/c)^k] \quad (1)$$

where V_j is defined by

$$V_j = [(P_j/P_r) - a]\, V_r/b \quad (2)$$

where P_r is the rated power of the wind turbine generator, V_r the rated speed of the wind turbine generator, and parameters a and b are empirical constants estimated by Justus [4] for the Pacific Coast region to be −0.42 and 1.14, respectively. In our numerical calculations we use data for machines characterized by P_r = 2000 kW, V_r = 10.6 m/s.

Equation (2) is just a linear approximation to the cubic power curve. The constants a and b will differ regionally and must be estimated from numerical simulation.

To use this model to test for the sensitivity of dispersal benefit to the region considered, we need an expression for the dependence of the array Weibull parameters on correlation coefficient. This is given by eqns. (3) and (4) below. Equation (3) expresses the relationship between the standard deviation of array wind speed frequency and both the standard deviation of the representative site speed distribution and the average array correlation coefficient:

$$\sigma_A = \{\sigma_T^2[1 + (n-1)\bar{\rho}]/n\}^{1/2} \tag{3}$$

where σ_A is the standard deviation of the array wind speed distribution, σ_T the standard deviation of the representative site wind speed distribution, $\bar{\rho}$ the average array wind speed cross-correlation coefficient, and n the number of sites in the array.

Equation (4) relates the Weibull shape factor k to σ and $\bar{v}$, the mean wind speed. This relation holds for both arrays and individual sites.

$$k = (\sigma/\bar{v})^{-1.086} \tag{4}$$

Equation (4) is an empirical approximation to the theoretical relationship between k and $\sigma/\bar{v}$ given by

$$(\sigma/v)^2 = [\Gamma(1 + 2/k)/\Gamma^2(1 + 1/k)] - 1$$

For the sake of completeness, we write down an expression for the Weibull scale factor c as a function of k and $\bar{v}$ as follows:

$$c = \gamma\bar{v}/\Gamma(1 + 1/k) \tag{5}$$

where γ is an empirical adjustment factor found to be about 1.02 - 1.03.

3.3. *Fitting the Weibull model to the Pacific Coast array*

We begin by using eqns. (1) and (2) to characterize the performance of Justus' Pacific Coast wind turbine array. We want to know if the simple model is a good approximation to numerical simulation of large arrays. In Table 4 we list data for summer

TABLE 4

Pacific Coast array — diurnal power variations: summer

		Hour			Data source
		13	16	19	
Mean power (kW)		840	1054	805	Table C-6 of ref. 4
Mean wind speed (m/s)		8.0	8.8	7.8	Table 7*
Weibull parameters by month					
July 73	c	9.21	10.11	9.29	
	k	3.13	3.77	3.53	
July 75	c		10.27	9.42	
	k		3.78	3.86	
August 71	c	9.00	9.85	8.61	
	k	3.52	3.99	3.67	
August 72	c	9.40	9.99	9.08	
	k	3.34	3.86	3.55	
August 74	c	9.21	9.68	8.77	
	k	3.13	3.78	2.87	
August 75	c	9.30	9.76	8.65	
	k	3.33	3.36	2.99	
Summer average	c	9.22	9.94	8.97	
	k	3.29	3.36	3.41	
$\Pr[x \leqslant \mu]$		0.47	0.47	0.46	
Median wind speed (estimated) (m/s)		8.3	8.9	8.1	

*C. G. Justus, personal communication.

diurnal variations in array output. The data on mean power and mean wind speed are simulated; they come from ref. 4. We also tabulate array Weibull parameters. These parameters used in eqns. (1) and (2) ought to be able to reproduce approximately the mean power and wind speed results from numerical calculation.

We make the test indirectly. If the Weibull distribution were symmetric, then the mean power would have probability equal to 0.50, *i.e.*, the mean and median coincide. Bury [11] shows that the Weibull approximation is symmetric for $k = 3.6$, which is quite close to our values. We can use a simple formula to calculate the median wind speed v_0; namely

$$v_0 = c(\ln 2)^{1/k} \qquad (6)$$

When eqn. (6) is substituted into eqn. (1) for V_j, then the cumulative probability will be 0.5, which is the definition of median.

Putting in our parameters we find median wind speeds that are 1 - 4% higher than the mean. Since our Weibull shape parameter k takes values less than 3.6, the distribution is skewed slightly to the right. Hence the median should be less than the mean, but only slightly [11]. Our data also show that the cumulative probability of the mean power ranges from 0.46 to 0.47. The combination of these errors could be as much as 10%. For our purposes later we shall rely principally on the 4:00 p.m. data, where the error is about 5%. This can give us reasonable confidence in the accuracy of (1) and (2), at least in this region.

It is more difficult to assess the accuracy of the Weibull model in the tail of the distribution. In some senses this is a more critical region for LOLP calculations. At this point we can only point out a potential difficulty in accuracy, but we do not have the tools to assess its import. Clearly this question can be answered by more simulation studies.

It is worth observing in passing that the diurnal variation data show rather large output for afternoon performance. The range 800 - 1000 kW per 2000 kW of capacity is 40 - 50% of rated capacity. This will be important in our capacity credit calculations. For now it remains to explore the use of eqns. (3) - (5) in the construction of various array power distributions. This is the subject to which we now turn.

3.4. California arrays: building up from the bottom and down from the top

In Fig. 1 we show two power output frequency distributions, one representing average conditions at Sacramento, California, the other representing the summer average performance of Justus' Pacific Coast array. It is useful in understanding the dispersal effect to build up arrays incrementally. This will illustrate how reliability increases and helps develop intuition for the tradeoffs involved. In the calculations which follow, attention is focused on summer conditions. The rationale for this limitation will be discussed more fully in § 4, but it is based primarily on the summer peaking electric demand behavior which is characteristic of California [12].

Our procedure is to use Sacramento as the "representative site" and gradually add additional sites to our arrays. From estimates developed by Justus *et al.* (ref. 8, pt. 21), we have Weibull parameters for Sacramento of $k = 1.98$ and $c = 6.32$. These have been scaled up to 60 m by empirical relations also developed in ref. 8. From eqn. (4) we can calculate a standard deviation of the wind speed distribution at this site. For array construction we use eqn. (3) and Table 3 to scale the array wind speed standard deviation. Then the data in Table 2 allow us to calculate $\bar{v}$, the mean array wind speed. This goes back into eqns. (4) and (5) to yield array Weibull parameters. Then eqns. (1) and (2) allow us to calculate a table of cumulative power frequency. This can be displayed graphically as in Fig. 1. The table corresponding to the single-site curve in Fig. 1, Sacramento, is given in Table 5.

TABLE 5

Sacramento: power output variations

$\Pr[P \leqslant 1000 \text{ kW}]$	= 0.84
$\Pr[P \leqslant 800 \text{ kW}]$	= 0.77
$\Pr[P \leqslant 600 \text{ kW}]$	= 0.67
$\Pr[P \leqslant 400 \text{ kW}]$	= 0.57
$\Pr[P \leqslant 200 \text{ kW}]$	= 0.44
$\Pr[P \leqslant 100 \text{ kW}]$	= 0.38

In Table 6 we tabulate average power frequency distributions for seven array configurations during the summer. In general, as the number of sites increases, so does reliability. We shall see that for utility system planning purposes, it is the behavior at lower power

TABLE 6

California sub-arrays: average summer power frequency distribution

Case 1		
Sacramento	$\bar{v}$ = 6.75 m/s	Pr[$P \leqslant$ 1000 kW] = 0.72
Stockton	$\bar{\rho}$ = 0.49	800 = 0.62
	k = 2.42	600 = 0.50
	c = 7.77	400 = 0.38
		100 = 0.22
		50 = 0.20
Case 2		
Bakersfield	$\bar{v}$ = 6.5 m/s	Pr[$P \leqslant$ 1000 kW] = 0.76
Stockton	$\bar{\rho}$ = 0.40	800 = 0.65
Sacramento	k = 2.61	600 = 0.53
	c = 7.46	400 = 0.40
		200 = 0.27
Case 3		
Stockton	$\bar{v}$ = 7.3 m/s	Pr[$P \leqslant$ 1000 kW] = 0.66
Sacramento	$\bar{\rho}$ = 0.41	800 = 0.53
San Francisco	k = 2.96	600 = 0.40
	c = 8.38	400 = 0.28
		200 = 0.18
Case 4		
Bakersfield	$\bar{v}$ = 7.0 m/s	Pr[$P \leqslant$ 1000 kW] = 0.71
Stockton	$\bar{\rho}$ = 0.39	800 = 0.58
Sacramento	k = 2.98	600 = 0.44
San Francisco	c = 8.00	400 = 0.31
		200 = 0.20
Case 5		
Red Bluff	$\bar{v}$ = 7.1 m/s	Pr[$P \leqslant$ 1000 kW] = 0.70
San Francisco	$\bar{\rho}$ = 0.34	800 = 0.56
Sacramento	k = 3.15	600 = 0.42
Stockton	c = 8.09	400 = 0.29
		200 = 0.18
Case 6		
Red Bluff	$\bar{v}$ = 6.9 m/s	Pr[$P \leqslant$ 1000 kW] = 0.73
Sacramento	$\bar{\rho}$ = 0.35	800 = 0.60
Stockton	k = 3.15	600 = 0.45
Bakersfield	c = 7.86	400 = 0.31
San Francisco		200 = 0.19
Case 7		
Pacific Coast array	$\bar{v}$ = 6.7 m/s	Pr[$P \leqslant$ 1000 kW] = 0.80
(13 sites)	$\bar{\rho}$ = 0.25	800 = 0.65
	k = 3.88	600 = 0.47
	c = 7.55	400 = 0.30
		200 = 0.16
		100 = 0.11

output levels which is crucial. In this regard it is interesting to pay particular attention to case 7, the entire Pacific Coast array. The mean summer wind speed over the whole array is less than that of most combinations of California sites. But the behavior at low output levels is better. This is due to low correlation among a significant number of sites.

Detailed examination of Table 6 shows the tradeoff between array mean wind speed $\bar{v}$ and average correlation coefficient $\bar{\rho}$. Consider cases 3 and 4. Case 3 involves three sites (Stockton, Sacramento and San Francisco). The power frequency distribution shows a little better reliability than the array composed of the same three sites plus Bakersfield, which is case 4. Adding Bakersfield to the case 3 array reduces both mean wind speed and average correlation coefficient. The former decreases by about 4%, the latter by about 5%. Looking at case 5 we find slightly better availability than case 3. In case 5 we add Red Bluff to the case 3 array instead of to Bakersfield. In this case mean wind speed falls only about 3%, and the correlation coefficient falls about 17%.

It would also be useful to characterize the California array performance with respect to diurnal variation in output as Table 4 does for the whole Pacific Coast array. To do this we must scale down from the large system to the smaller ones of interest. We cannot use the "building up" approach which we used for our summer average calculations because there are no base case data for a single representative site. Nonetheless eqns. (1) - (5) are sufficiently flexible for a simple scaling procedure to be possible. From eqn. (3) we know the ratio of an array wind speed standard deviation to that of a representative site. This is expressed in terms of the pair of numbers $(n, \bar{\rho})$, where n is the number of array sites and $\bar{\rho}$ is the average correlation coefficient. We calculate this ratio for the Pacific Coast array (n = 13, $\bar{\rho}$ = 0.25) and get the value 0.555. Doing the same for case 2 (n = 5, $\bar{\rho}$ = 0.35) we get the value 0.693. Therefore, we know that

$$\sigma(\text{PG\&E})/\alpha(\text{Pacific Coast array}) = 0.693/0.555 = 1.249$$

We calculate σ(Pacific Coast array) from eqn. (4) using the known values of k = 3.76 and $\bar{v}$ = 8.8 m/s. This gives σ(Pacific Coast array) = 2.60, and therefore k, using eqn. (4) again, but this time reestimating $\bar{v}$ for the sites in question. This is done by averaging the mean values in Table 2. For PG&E data this yields $\bar{v}$ = 6.86 compared with the array average of 6.7 m/s. Using a linear scaling ratio we get a 4:00 p.m. $\bar{v}$ = (6.86/6.7)8.8 = 9.0 m/s. Now returning to eqn. (4) we have $k = (9.0/3.246)^{1.086}$ = 3.03. From eqn. (5) we get c = 10.28.

A similar procedure is used to derive the parameters for case 3. For this case n = 6, $\bar{\rho}$ = 0.27. The latter value is calculated from Table 3. Results are summarized in Table 7.

A summary of selected distributions is shown in Fig. 3. The next problem is to decide how to use the abundance of data for capacity credit calculations.

4. CAPACITY CREDIT FOR WIND ARRAYS: FORMULATING THE PROBLEM

4.1. *The coincidence hypothesis*

Our first problem in array analysis is the selection of the appropriate data. Do we use

TABLE 7

Power frequency distribution: 4:00 p.m. summer

Case 1			
Pacific Coast array	$\bar{v}$ = 8.8 m/s	Pr[$P \leq$ 1400 kW]	= 0.70
(Table 4)	k = 3.76	1200	= 0.57
	c = 9.94	1000	= 0.43
		800	= 0.31
		600	= 0.20
		400	= 0.12
		200	= 0.06
Case 2			
Pacific Gas and	$\bar{v}$ = 9.0 m/s	Pr[$P \leq$ 1000 kW]	= 0.44
Electric Company	k = 3.03	800	= 0.33
sites	c = 10.28	600	= 0.24
		400	= 0.16
		200	= 0.10
Case 3			
All California sites	$\bar{v}$ = 9.17 m/s	Pr[$P \leq$ 1000 kW]	= 0.40
	k = 3.45	800	= 0.29
	c = 10.40	600	= 0.20
		400	= 0.12
		200	= 0.07

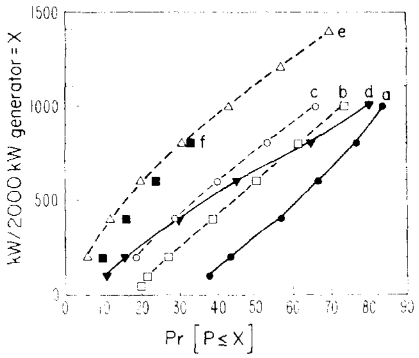

Fig. 3. Power frequency distribution: various configurations. Curve a, Sacramento (Table 5); curve b, Sacramento and Stockton (Table 6, case 1); curve c, Sacramento, Stockton and San Francisco (Table 6, case 3); curve d, Pacific Coast array summer average (Table 6, case 7); curve e, Pacific Coast array, 4 p.m. summer (Table 7, case 1); curve f, PG&E, 4 p.m. summer (Table 7, case 2).

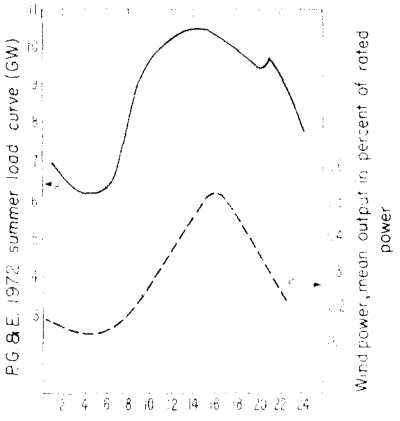

Fig. 4. Coincidence of demand and wind resource.

averages? What is the limitation of restricting attention to the summer period? What guidelines are there to help us avoid just simulating all available data with brute force methods? The principle we shall follow is essentially the conventional wisdom in power system studies. What matters for power system reliability is the coincidence in the power availability and the peak demand. Let us call this principle the "coincidence hypothesis".

The coincidence hypothesis is more than just folklore. Recent studies of other intermittent electric technologies have verified the importance of this measure. General Electric Company, in a study of photovoltaic power plants, found that degree of coincidence in peak load and insolation was a significant parameter [13]. This coincidence requirement refers to hourly correspondences. Seasonal coincidence is assumed. This is the rationale for our looking at summer data only. But the hourly scale requires that we focus on diurnal variation. Therefore, the daily average data of Table 6 is too coarse a measure. For our purposes the best approximation is the data given in Table 7. The summer time peak in California is typically around 3:00 p.m. [12]. The diurnal variations in array output are compared with the demand curve of the Pacific Gas and Electric Company in Fig. 4. This shows good coincidence with our 4:00 p.m. data.

This choice of power frequency distribution is reasonable provided we believe that significant mismatch between loads and resources would not emerge at some other time. There are a number of potential problems here. For example, Fig. 4 shows a growing gap between 9:00 p.m. and midnight. This mismatch is probably not inherent in the wind regime and could be ameliorated with more sophisticated siting. Data for the San Bernardino Mountains, for example, show virtually an antisymmetric wind pattern to Fig. 4 with a minimum at 4:00 p.m., increasing in the evening [14]. Moreover, General Electric in another study [15] suggests that high wind speed regimes typically have a diurnal pattern which complements the lower wind speed pattern we have illustrated in Fig. 2. Figure 5 shows this change as average

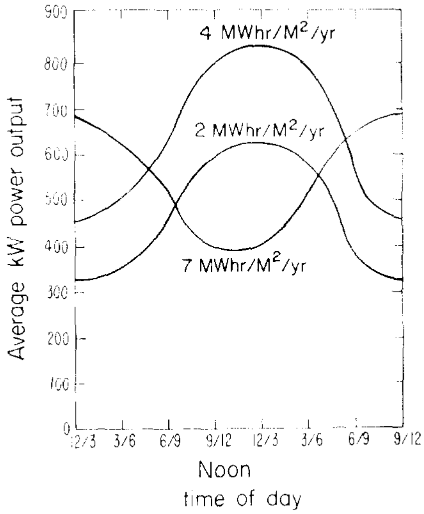

Fig. 5. Diurnal power output variations for three wind regimes.

energy increases. Thus, for the purposes of illustration, it is not unreasonable to accept the 4:00 p.m. power frequency distribution as the basic data to be used in analysis of summer conditions.

An interesting potential problem is the emergence of a winter peak shortfall. Table 2 shows that the northern California sites have a somewhat lower mean wind speed in winter compared with that in summer (5.6 m/s *vs.* 6.9 m/s). An estimate of when a problem might emerge can be derived from a comparison of the summer/winter peak load differential and the effective load-carrying capability of a wind array. In § 4.3 below we shall present results on ELCC for various penetrations of wind turbines. If these values of ELCC were greater than the difference between summer and winter peak loads, then poorer performance in winter would threaten reliability at the winter peak. For the Pacific Gas and Electric Company service area, the winter peak is typically more than 2000 MW below the summer peak. Our results in § 4.3 will show that ELCC in this system will not be likely to exceed such a number. Therefore winter risks would not emerge under our assumptions. Of course, the question deserves more analysis than these brief remarks. But that is beyond the scope of the present effort.

4.2. A map from power frequency to generator models

We have argued briefly in § 2.1 that the translation of a wind power frequency distribution into a representation for LOLP analysis is non-trivial. The basic problem is that the power behavior of wind generators cannot be captured in two or three discrete states which is the typical model for thermal units. Indeed, it has been shown that even for thermal units LOLP accuracy is improved by using a multi-state model rather than collapsing behavior into an "on-off" model [3, 16]. The general rule here is that accuracy improves as the number of states increases. There is, of course, a point of diminishing returns. To limit computational complexity we shall use an eight-state model which is described below. Sensitivity of results was studied using other output models. These calculations support the accuracy of our model. Details are reported in Appendix 2 of ref. 10.

Before describing our transformation from a continuous distribution to the eight-state model, it is worth emphasizing that we need to model the low output end of the distribution more finely than the high output side. This thesis is also supported by data discussed in Appendix 2 of ref. 10. However, there is an intuitive rationale for our procedure that is useful to discuss. LOLP, our measure of risk and reliability, is naturally more sensitive to big failures than to small ones. The risk of using wind power for electricity is the risk of lulls in the wind. Even slight winds are better than no winds at all. The thesis which states that low output is critical captures the importance of marginal winds. At a time of risk anything is a whole lot better than nothing.

Let us consider case 2 from Table 7. We shall select eight states from the infinite possibilities and use interval estimates to characterize the probability of the state at the bottom of the interval. For example, if we use the case 2, Table 7 data and look at the interval between 200 and 400 kW, we find that the probability of power in this interval is 0.06. This is just the difference between $\Pr[P \leqslant 400]$ and $\Pr[P \leqslant 200]$ or $0.16 - 0.10 = 0.06$. This rule is conservative. It throws out all output between 200 and 400 kW and counts it all as just 200 kW. In Table 8 we show our generator model for this case.

A final note should be addressed to the question of the penetration of the wind array in the power system. In our analysis we scale the generator representation linearly with penetration. Thus a 2000 MW penetration in PG&E would be modeled by the states listed in Table 8 where each state had units of

TABLE 8

PG&E 4:00 p.m. summer wind generator model

State (X_i)	$\Pr[P \leqslant X_i]$	$\text{Prob}[\text{State } X_i] = \Pr[X_i \leqslant P \leqslant X_{i+1}]$
0	0.05	0.07
100	0.07	0.03
200	0.10	0.06
400	0.16	0.08
600	0.24	0.09
800	0.33	0.11
1000	0.44	0.30
1600	0.74	0.26

megawatts. For a 4000 MW penetration each state would have units of twice that number of megawatts. As the penetration grows the gap between states gets larger, that is, the grain of the representation gets coarser. This plus the linear scaling assumption may introduce problems that require further research.

4.3. Results from LOLP calculations

We have asked the following question about California wind arrays. Does the better reliability of the statewide wind resource (Table 7, case 3) compared with that of the PG&E area (Table 7, case 2) mean more capacity credit for statewide implementation? To answer this question we first must calculate the ELCC. We use a standard LOLP model [4] and data on the characteristics of the California electric generation system [17]. The results of these calculations are presented in Table 9.

The data in Table 9 do not provide an unambiguous answer to our question. Clearly the hypothetical California pool can carry more load with wind turbines than the PG&E service area alone. However, the ratio of ELCC to wind turbine capacity is better for PG&E. This measure is a better indicator of effectiveness than just ELCC. The results suggest the conclusion we began with; namely, that for these data the diversity benefit reaches diminishing returns when we go beyond northern California. This result is not conclusive, of course. Better array design, more and better wind correlation data, and so on could change the outcome. The methodological conclusion of interest is that even when wind output reliability is better in one region, ELCC is dependent on the characteristics of the systems in question.

4.4. Capacity credit

Capacity credit in our case means the amount of conventional capacity which can be displaced by wind generation. Our ELCC calculations form the basis for making this comparison. What we need to know is the ELCC of competing alternatives. As should be clear by now ELCC is a function of unit size, forced outage rate and the power system in question. We shall consider 800 MW coal plants in the PG&E system first, since these are being proposed.

A useful tool for evaluating ELCC where some simulation has already been done is the linear approximation to LOLP introduced by Garver in ref. 18. This technique has achieved widespread acceptance in the power industry. Garver derives the following expression:

$$\mathrm{ELCC} = C - m \ln[(1-r) + re^{c/m}] \qquad (7)$$

where C is the nominal generator rating, m the characteristic slope of the LOLP curve, and r the generator forced outage rate. The main parameter to be estimated in this equation is the slope of the risk (LOLP) curve.

To calculate m we look at a curve of LOLP *vs.* load and find the difference in megawatts between the risk criterion (one day in 10 years) and e = 2.718 times that risk. Recalling that one day in 10 years has the interpretation of 8.5×10^{-3} (see Appendix 2, ref. 10), it was found that m = 260 for the base PG&E system. Now we can use eqn. (7) to calculate ELCC for an 800 MW unit on the PG&E

TABLE 9

Wind array ELCC

PG&E			
(1) Wind turbine capacity (MW)	2000	4000	5000
(2) ELCC (MW)	530	700	740
(3) Base hydro-thermal capacity = 16 354 MW			
(4) Penetration (=(1)/[(1) + (3)])	11%	20%	23%
(5) Ratio of (2) to (1)	0.265	0.175	0.148
California utilities pooled			
(1) Wind turbine capacity (MW)	4000	8000	10000
(2) ELCC (MW)	970	1190	1240
(3) Base hydro-thermal capacity = 39 492 MW			
(4) Penetration (=(1)/[(1) + (3)])	9%	17%	20%
(5) Ratio of (2) to (1)	0.243	0.149	0.124

system as a function of forecasted forced outage rate. PG&E forecasts a 0.12 forced outage rate for such units when they mature (after 4 years) [14]. Using that value, the ELCC would be about 425 MW. The Electrical Power Research Institute, however, estimates a forced outage rate (FOR) of 0.197 [19]. At that level of performance, the ELCC would be approximately 380 MW. Table 10 shows the ELCC corresponding to each. An intermediate value of 0.15 yields an ELCC of 430.

TABLE 10

ELCC for PG&E 800 MW unit

FOR	ELCC
0.12	475
0.15	430
0.197	380

These results change when we look at a hypothetical pool of all California utilities. For this purpose let us consider a 1000 MW plant with a forced outage rate of 0.15. From ref. 17, the parameter m can be calculated to be 400. Then, using eqn. (7), we find that such a unit would have an ELCC equal to about 606 MW. Two such units would be displaced by about 8000 MW of wind turbine capacity. The ratio of wind generator capacity to conventional capacity displaced is thus about four to one in this case. That compares with something less than three to one in the PG&E case. This is another way of seeing the diminishing returns to implementing statewide pooling of the wind resource. Perhaps with better data this conclusion would change.

Finally, it should be noticed that Table 9 illustrates the barrier of large penetrations. The PG&E data show an increase of only 40 MW in ELCC for the last 1000 MW of wind turbine capacity. On a relative basis the situation is worse for the statewide case. These results suggest that we are near a limit. It is not clear if the limit is inherent in the resource or only in our limited data. Even if it is the latter, it is likely that this barrier will remain; only its frontier will recede.

5. CONCLUSIONS

Wind generators can displace conventional capacity with the reliability that has been traditional in power systems. The fraction of rated wind turbine capacity that can carry load reliability will decrease with increasing penetration of this resource in the generator mix. The geographical averaging of wind power output also shows diminishing marginal returns as the region under consideration increases. Even regions with highly reliable wind regimes may not show the same proportional capacity credit as smaller regions with less reliable wind because of the influence of the base generator mix.

6. THE RESEARCH HORIZON

Modeling the reliability of wind arrays is still a primitive subject. There is a great need for additional wind resource data. We need to know the time variation of correlations, more details on different wind regimes, and the influence of differing machine parameters. Even with our present scanty knowledge, it seems plausible to assert that our main thesis has been supported. Wind arrays can displace conventional capacity with reliability.

The present inquiry forms part of a larger subject, the integration of intermittent resource technologies into an electric energy system. Our discussion has focused only on the problem of relatively small penetrations. Large penetration analysis will require a look at hydro optimization, other resource complementarities, and the role of storage. This is a large and fruitful area of study which will help determine the future economic viability of emerging technologies.

ACKNOWLEDGEMENTS

This work was supported by the UC-LBL Distributed Technology Project and the California Energy Commission. I would like to thank Paul Craig, Mark Levine, Meir Carasso and Kish Sharma for their encouragement. An earlier version of this report was considerably improved by the comments of Richard Sextro (who also provided computa-

tional help), Carl Blumstein, C. G. Justus, Amory Lovins and Neil Woodley. The responsibility for remaining errors is the author's.

REFERENCES

1 Stanford Research Institute International, A Comparative Evaluation of Solar Alternatives: Implications for Federal RD&D, Vol. 1, Menlo Park, CA, January 1978.

2 J. P. Molly, Balancing power supply from wind energy converting systems, Wind Engineering, 1 (1977) 57 - 66.

3 E. Kahn, Testimony before New Jersey PUC, Construction Docket 762-194, filed June 1977.

4 C. G. Justus and W. R. Hargraves, Wind Energy Statistics for Large Array of Wind Turbines (Great Lakes and Pacific Coast Regions), Georgia Institute of Technology, RLO/2439-77/2, May 1977.

5 R. Sullivan, Power System Planning, McGraw-Hill, New York, 1977.

6 E. Kahn, An Assessment of the Potential for Full Coordination of the California Electric Utilities, LBL-5941, January 1977.

7 K. Sheehy, Legal and Institutional Factors Affecting the Implementation of Power Pooling Within California, CERCDC, 1978.

8 C. G. Justus, W. R. Hargraves and Amir Mikhail, Reference Wind Speed Distributions and Height Profiles for Wind Turbine Design and Performance Evaluation Applications, Georgia Institute of Technology, ORO/5108-76/4, August 1976.

9 W. Cliff, The Effect of Generalized Wind Characteristics on Annual Power Estimates from Wind Turbine Generators, PNL-2436, October 1977.

10 E. Kahn, Reliability of Wind Power From Dispersed Sites: A Preliminary Assessment, Lawrence Berkeley Laboratory, Berkeley, LBL - 6889, April, 1978.

11 K. Bury, Statistical Models in Applied Science, Wiley, New York, 1975.

12 D. E. Morse, Technical Documentation of Staff Procedures for Establishing Peak Demand, ERCDC, October 1976.

13 General Electric Company, Requirements Assessment of Photovoltaic Power in Electric Utility Systems, EPRI RP 651-1, Summary Report, February 1978.

14 Pacific Gas and Electric Company, Fossil 1 and 2 NOI, December 1977.

15 General Electric Company, Wind Energy Mission Analysis, C00/2578-1/2, February 17, 1977.

16 R. A. Billinton, V. Jain and C. MacGowan, Effect of partial outage representation in generation system planning studies. IEEE Trans., PAS-93 (1974) 1252-59.

17 J. Sathaye *et al.*, Potential Electricity Impacts of a 1978 California Drought, Lawrence Berkeley Laboratory, Berkeley, LBL-6871, January 1978.

18 L. L. Garver, Effective load carrying capability of generating units. IEEE Trans., PAS-85 (1966) 910 - 19.

19 Electrical Power Research Institute, Technical Assessment Guide, EPRI, Palo Alto, August 1977.