Tractable Analytic Expressions for the Wind Speed Probability Density Functions Using Expansions of Orthogonal Polynomials

MARK L. MORRISSEY

School of Meteorology, University of Oklahoma, Norman, Oklahoma

J. SCOTT GREENE

Department of Geography and Environmental Sustainability, University of Oklahoma, Norman, Oklahoma

(Manuscript received 21 November 2011, in final form 23 February 2012)

ABSTRACT

The use of the two-parameter Weibull function as an estimator of the wind speed probability density function (PDF) is known to be problematic when a high accuracy of fit is required, such as in the computation of the wind power density function. Various types of nonparametric kernels can provide excellent fits to wind speed histograms but cannot provide tractable analytical expressions. Analytic expressions for the wind speed PDF are needed for many applications, particularly in the downscaling of model or satellite wind speed estimates to the regional or point scale. It is demonstrated that the judicious use of an expansion of orthogonal polynomials can produce more accurate estimates of the wind speed PDF than relatively simply parametric functions, such as the commonly used Weibull function. This study examines four such expansions applied to two different surface wind speed datasets in Oklahoma. The results indicate that the accuracy of fit of a given expansion is strongly related to how close the basis weight function in an expansion resembles the wind speed histogram. It is shown that this basis function, which is the first term in the expansion, acts as a first "best guess" to the true wind speed PDF and that the additional terms act to "adjust" the fit to converge on the true density function. The results indicate that appropriately chosen orthogonal polynomials can provide an excellent fit and are quite tractable.

1. Introduction

The wind speed probability density function (PDF) is a crucial element in many scientific and engineering investigations. For example, it is used by the wind power industry to assess turbine blade fatigue (Veers 1983; White 2004) and to determine the wind speed threshold for turbine cutoff and cutin (Duenas-Osorio and Basu 2008). Most important, it is used to compute the wind speed power density (WPD) to assess the wind power potential at specific locations (Li and Li 2005; Lackner et al. 2008; Morrissey et al. 2010b). For general meteorological applications, it is often used to parameterize surface energy fluxes and momentum in general circulation models (GCMs) since a bias in the estimation of these fluxes can occur if grid-scale GCM wind speed is used in their calculation (Vickers and Esbensen 1998; Monahan 2006a; Monahan 2007). The bias results from the nonlinearity in the relationship between wind speed and the surface fluxes and momentum (Wright and Thompson 1983; Wang et al. 1998; Feely et al. 2004). Capps and Zender (2008) demonstrated that GCM predictions improve with the inclusion of subgrid-scale wind speed PDFs in the models instead of the grid-scale mean. In a study of dust emissions simulated with a general circulation model, Cakmur and Miller (2004) modeled the subgrid-scale wind speed variability by assuming rather simple Weibull and bivariate normal parametric wind speed PDFs where the function parameters were estimated from model parameterizations of the boundary layer processes. The use of the wind speed PDF was predicated by knowledge that dust emission occurs once the wind speed exceeds a given threshold whose probability of occurrence can be found from the PDF. In their paper, they comment that their work may benefit from the application of PDF functions with improved accuracy.

Corresponding author address: Mark L. Morrissey, School of Meteorology, University of Oklahoma, Ste. 5900, 120 David L. Boren Blvd., Norman, OK 73026.
E-mail: mmorriss@ou.edu

DOI: 10.1175/JAMC-D-11-0236.1

The two-parameter Weibull distribution has generally been accepted as an adequate model for the wind speed PDF (Hennessey 1977; Justus et al. 1978; Pavia and O'Brien 1986; Barthelmie and Pryor 2003; Pryor et al. 2004; Ramirez and Carta 2005; Monahan 2006b). However, wind speed does not always have a Weibull-like distribution (Jaramillo and Borja 2004; Yilmaz and Çelik 2008; Morrissey et al. 2010b). In addition, the shape of the PDF depends upon the measurement scale (Morrissey et al. 2010a). In some instances, the Rayleigh (Baker and Hennessey 1977; Hennessey 1977) or lognormal models (Justus et al. 1978) have been used. For WPD computations, large errors can result from even minor errors in the estimation of the true wind speed PDF since the WPD is a function of the expected value of the cube of the wind speed.

Excellent fits to wind speed histograms can be made using a nonparametric kernel approach (Silverman 1998; Juban et al. 2007). As an example, Monahan et al. (2011) used kernel density estimates of the joint PDFs of near-surface and 200-m winds to examine the relationship between the winds at these two levels. But, as noted by Bryukhan and Diab (1993) and Morrissey et al. (2010b), many applied climate applications require a tractable analytic expression for the wind speed PDF that cannot be obtained from the kernel method (Silverman 1998), as the number of terms in kernel expressions equals the number of data points used in the fit. A tractable analytic expression for the wind speed density function is especially important for downscaling either wind speed PDFs or wind speed itself from model and satellite estimates (Pryor et al. 2005). A common approach to this problem is to determine a statistical (Mearns et al. 1999; Pryor et al. 2005) or dynamical (Zagar et al. 2006; Frech et al. 2007; Gustafson and Leung 2007) relationship between certain model output variables and the parameters of a wind speed PDF. If an analytic expression that provides an accurate fit to the wind speed PDF can be clearly identified, it is difficult to see how nonparametric kernel estimators could provide more useful, parsimonious functions for use in downscaling and related problems.

One method of interest that potentially meets the above requirements for a tractable wind speed PDF estimator is an expansion of orthogonal polynomials. While not a new concept, there are few examples of the use of such expansions in the meteorological literature. Two exceptions include the use of the classical Hermite orthogonal polynomials by Monahan (2006a) and Morrissey et al. (2010a). In the wind engineering field, Bryukhan and Diab (1993) developed an expansion for the wind speed PDF using a different function: the classical Laguerre polynomial. While such expansions can provide tractable analytic expressions for the wind speed PDF, the aforementioned studies identified the many uncertainties as to which expansion provides the ''best'' fit with smallest number of terms in the expansion. In addition, polynomial expansions can prove troublesome with the possibility that the resulting functions can occasionally give negative density values, which are clearly impossible to obtain. Also, there are seemingly different ways to expand the same polynomial. For example, Monahan (2006a) used a Gram–Charlier series A expansion of Hermite polynomials (Johnson et al. 1994) to fit ocean wind speed PDFs while Morrissey et al. (2010a) used a Gauss–Hermite expansion to estimate the wind WPD. Working with galaxy spectral lines, Blinnikov and Moessner (1998) demonstrated that the Gauss–Hermite expansion has much better convergence properties than did the Gram–Charlier Series A expansion near-Gaussian-shaped distributions. A review of the above studies strongly suggests that the optimal expansion method and type of orthogonal polynomial used is a function of the analytical nature of the underlying shape of the wind speed PDF, which is unknown, but can be hypothesized using a histogram. What is needed is a general approach to fitting orthogonal polynomial expansions whereby the data determine the type of orthogonal polynomial and expansion used. One logical approach, put forth by Provost and Jiang (2012), is to first identify a best-fit parametric base density and then to use this density as the weight function and, thus, as the first term in a polynomial expansion. For example, if the wind speed histogram has a Gaussian-like shape, then a base function proportion to e^{-V^2} (V being the random variable wind speed) would be used. Likewise, an exponentially shaped histogram would suggest the use of a weight function proportional to e^{-V}.

The works of Provost (2005) and Provost and Jiang (2012) show that, given an appropriate basis weight function, the addition of orthogonalized polynomials in the expansion can be used to adjust the fit until certain smoothing criteria are met and an acceptable PDF estimator is obtained. The purpose of this paper is to examine and compare four approaches proposed in the literature: 1) the Laguerre expansion by Bryukhan and Diab (1993); 2) the Gauss–Hermite expansion by Morrissey et al. (2010a), 3) a ''generalized'' Laguerre polynomial expansion (Provost and Jiang 2012), and 4) a method whereby any set of orthogonal polynomials can be constructed from a given weight function. With the latter method, which uses the Gram–Schmdit orthogonalization process (Afken and Weber 2005), the Weibull PDF will be used as the basis weighting function in the expansion of these polynomials. The Laguerre and Gauss–Hermite expansions formulized by Bryukhan and Diab (1993) and

Morrissey et al. (2010b), respectively, are described first. The generalized Laguerre polynomial scheme is then given. Each formulation is developed in a similar fashion. Finally, the general approach using the Gram–Schmidt method incorporating a Weibull PDF weight is shown. The mathematical similarities and differences among these developments will then be summarized and discussed. Two near-surface (10 m) wind speed datasets from two locations obtained from the Oklahoma Mesonetwork are used to examine the usefulness of each expansion. The mesonet uses an R. M. Young Company 5103 anemometer, which has a range of 1–60 m s^{-1} and which can withstand gusts of 100 m s^{-1}. A wind speed of 1 m s^{-1} is necessary to start the propeller. In-depth details concerning the wind sensors and their accuracy are given by Brock et al. (1995). The two locations were carefully selected so that the approaches described herein could be examined on wind speed histograms that have completely different shapes not captured by the two-parameter Weibull PDF function.

2. Method development

a. The Laguerre series expansion

Bryukhan and Diab (1993) noted that since wind speed is defined on the positive real line $[0, \infty)$ and generally has an exponential-like PDF, it seemed natural to use a Laguerre polynomial expansion, as this expansion has an exponential weight function and is also defined on the same support interval as the data. Bryukhan and Diab (1993) used a specific, simplified form of the Laguerre polynomial. A more "generalized" Laguerre polynomial has an additional parameter that gives it additional flexibility. Unfortunately, there is considerable discrepancy in the naming convention for the more general Laguerre polynomial in the literature. This paper follows the naming structure set forth by Szego's (2003, p. 100) work and will refer to the more general form as the generalized Laguerre polynomial and the more specific form simply as the Laguerre polynomial.

Bryukhan and Diab's (1993) formulation is as follows. Assume that wind speed measured at a location fixed in space is a random variable V with real-valued positive realizations υ, $[0 < V < \infty)$. For the Laguerre expansion, the realizations of the random variable are first standardized using an affine correction, $U = V/\sigma$, where σ is the standard deviation of the sample realizations of V, that is, υ. Thus, the random variable, U, has realizations, u, with a standard deviation of 1 m s^{-1}. The actual PDF, $f(V)$, can be approximated using an expansion truncated to $K + 1$ terms using

$$f(V) \sim f_{\mathrm{LE}}(V\,|\,K) = \frac{1}{\sigma}\sum_{k=0}^{K}\psi_k e^{-u}L_k(u),$$
$$u = \frac{\upsilon}{\sigma} \quad \forall \quad \{u \in R, 0 \le u < \infty\}, \qquad (1)$$

where $L_k(u)$ is a Laguerre polynomial of order k and ψ_k is the kth coefficient of the expansion. The weight function, e^{-u}, is the standardized exponential PDF with the variable u having a mean and standard deviation of 1 m s^{-1}. Applying the affine transform to the data is necessary so that the standardized variable, u in this case, has the first two moments of the standardized exponential weight function (Morrissey et al. 2010b; Provost and Jiang 2012). Laguerre polynomials can be constructed from the recursion relation,

$$L_0(u) = 1$$
$$L_1(u) = 1 - u$$
$$L_{k+1}(u) = \frac{1}{k+1}[(2k + 1 - u)L_k(u) - kL_{k-1}(u)], \qquad (2)$$

and, for any positive, real random variable X, they are orthonormal with respect to the weight function e^{-X}; that is,

$$\int_0^\infty e^{-x}L_k(x)L_j(x)\,dx = \begin{pmatrix} 1, & k = j \\ 0, & k \ne j \end{pmatrix}. \qquad (3)$$

By multiplying both sides of (1) by $L_k(\upsilon)$, integrating over the interval $[0, \infty)$, and using the orthonormal relation above, the expansion coefficients can be estimated from data for a given k value using

$$\psi_k = \sigma\int_0^\infty L_k(u)f_L(\upsilon)\,d\upsilon$$
$$= \int_0^\infty L_k(u)f_L(\upsilon)\,du \sim \frac{1}{n}\sum_{i=1}^{n}L_k[u(i)]. \qquad (4)$$

Note that the coefficients are obtained by estimating $E[L_k(u)]$ using n transformed data values, $u(i), i = 1, 2, \ldots, n$. This amounts to equating the powers of u in the polynomial $L_k(u)$ to the sample moments obtained from the data.

b. Gauss–Hermite series expansion

Hermite polynomials of order k—that is, $H_k(V)$—used in the Gauss–Hermite expansion are orthogonal with respect to the weighting function $e^{-\upsilon^2}$ for any real, continuous random variable V with realizations υ and are orthogonal on the interval $(-\infty < V < \infty)$. The Hermite polynomials defined for the Gauss–Hermite expansion can be generated using the recursion formula:

$$H_0(v) = 1$$
$$H_1(v) = 2v$$
$$H_{k+1}(v) = 2vH_k(v) - 2kH_{k-1}(v). \quad (5)$$

From Morrissey et al. (2010b), the wind speed PDF of V is approximated using a truncated (i.e., $K + 1$ terms) Gauss–Hermite expansion after the variable transformation to u by

$$f(V) \sim f_{\mathrm{GH}}(V \mid K) = \frac{1}{\sigma}\sum_{k=0}^{K}\psi_k H_k(u)g(u), \quad u = \frac{v-\mu}{\sigma} \quad \forall \quad \{u \in R, -\infty \le u \le \infty\}, \quad (6)$$

where $g(u)$ is the standard Gaussian PDF,

$$g(u) = \frac{1}{\sqrt{2\pi}}e^{-(u^2/2)}, \quad (7)$$

and u represents standardized wind speed values. Note that, as with the Laguerre expansion, the data are again affine transformed to have a mean and standard deviation commensurate with the weight function, which is a standard normal PDF (i.e., $\mu = 0$ m s^{-1}; $\sigma = 1.0$ m s^{-1}). Thus, in the same manner as the Laguerre expansion above, it is hypothesized that the data should have the same moments as the basis weight function. Hermite polynomials for a random variable X defined on the interval $(-\infty, \infty)$ have the following orthogonal relationship:

$$\frac{\sqrt{\pi}}{2^{k-1}k!}\int_{-\infty}^{\infty} H_k(x)H_j(x)g(x)^2\,dx = \begin{pmatrix} 1 & k = j \\ 0 & k \neq j \end{pmatrix}. \quad (8)$$

Following the same procedure as with the Laguerre expansion, the orthogonality property of Hermite polynomials allows the coefficients ψ_k of order k to be estimated from the data. First, we multiply both sides of (6) by $[\sqrt{\pi}/(2k^{-1}k!)]g(u)H_k(u)$ and integrate over the interval $(-\infty, \infty)$. This provides an expression for the expansion coefficients:

$$\psi_k = \frac{\sigma\sqrt{\pi}}{2^{k-1}k!}\int_{-\infty}^{\infty} f(v)H_k(u)g(u)\,dv = \frac{\sqrt{\pi}}{2^{k-1}k!}\int_{-\infty}^{\infty} f(v)H_k(u)g(u)\,du \sim \frac{\sqrt{\pi}}{2^{k-1}k!}\frac{1}{n}\sum_{i=1}^{n} H_k[u(i)]g[u(i)]. \quad (9)$$

Since the above equation represents a constant times the expected value of $H_k(u)g(u)$ for a given value of k, the kth expansion coefficient can be estimated from the dataset of size n.

c. Generalized Laguerre polynomials

In a manner similar to the above developments, the generalized Laguerre polynomial can also be used in an expansion whereby the data are again used to estimate the expansion coefficients. Since generalized Laguerre polynomials have an additional parameter, the deviation of the expansion and the expression for the coefficients are more involved than the two expansions above. We define a random variable V whose support is on the interval $[0, \infty)$. A scaled version of V is again necessary and takes the form of an affine correction $U = V/c$. It will be shown that the value c is obtained from equating the first two moments of the data to the basis weight function, which is a gamma PDF.

The moments of the transformed variable U are defined using

$$\mu_U(m) = E\left[\left(\frac{V}{c}\right)^m\right]. \quad (10)$$

The truncated expansion with generalized Laguerre polynomials with $K + 1$ terms takes the form

$$f(V \mid K) \sim f_{\mathrm{GL}}(V \mid K) = \frac{u^\tau e^{-u}}{c}\sum_{k=0}^{K}\psi_k L_k^\tau(u), \quad u = \frac{v}{c}, \quad (11)$$

where, from Szego (2003), the generalized Laguerre polynomials with parameter τ are defined by

$$L_k^\tau(u) = \sum_{i=0}^{k}(-1)^i\frac{\Gamma(k+\tau+1)u^{k-i}}{i!(k-i)!\Gamma(k+\tau-i+1)} \quad (12)$$

with $\Gamma(K + \tau + 1)$ being the gamma function and k the order of the polynomial. It should be noted that $L_k^0(u) = L_k(u)$ (Szego 2003). For any positive, real random variable X, the orthogonality relationship takes the form

$$\int_0^\infty x^\tau e^{-x}L_k^\tau(x)L_j^\tau(x)\,dx = \begin{bmatrix} \dfrac{\Gamma(k+\tau+1)}{k!} & k = j \\ 0 & k \neq j \end{bmatrix}. \quad (13)$$

To obtain the expansion coefficients, both sides of (11) are multiplied by $L_j^\tau(u)$ and integrated over the support interval $[0, \infty)$ to produce

$$c\int_0^\infty L_k^\tau(u)\,f_{\mathrm{GL}}(v)\,dv = \sum_{k=0}^{\infty}\psi_k\int_0^\infty u^\tau e^{-u}L_k^\tau(u)L_j^\tau(u)\,du, \quad (14)$$

noting that $du = dv/c$. For a given value of k with $k = j$ and using the orthogonality relationship an expression for the expansions coefficients is obtained:

$$\psi_k = \frac{k!}{\Gamma(k + \tau + 1)}\int_0^\infty L_k^\tau(u)f(v)\,dv. \qquad (15)$$

Employing Eq. (15), an estimator can be constructed using the fact that the integral represents the expected value of $L_k^\tau(u)$. This estimator can take the form

$$\psi_k \sim \frac{k!}{\Gamma(k + \tau + 1)}\frac{1}{n}\sum_{i=1}^{n} L_k^\tau[u(i)], \qquad (16)$$

with n being the sample size of the data. Provost (2005) makes an important insight that the expansion in (11) uses a gamma PDF as the first term with succeeding terms as ‘‘adjustments’’ to the density approximator to account for deviations of the true density from the gamma PDF. This is proven below. Thus, the problem now becomes how to get the first term in the expansion in the form of the required gamma PDF. It turns out that this can be accomplished by selecting certain expressions for c and τ in terms of the first two moments of the gamma PDF.

For a random variable X, the two-parameter gamma PDF for a random variable X takes the form

$$f_g(X\,|\,\alpha,\beta) = x^{\alpha-1}\frac{e^{-(x/\beta)}}{\Gamma(\alpha)\beta^\alpha}, \qquad (17)$$

with the first central moment being $\alpha\beta$ and the second central moment being $\alpha\beta^2$. Given that $L_0^\tau(u) = 1$, the first term in expansion (11), $k = 0$, given a change of variables to v (i.e., $v = uc$), is

$$\frac{e^{-(v/c)}v^\tau c^{-\tau}}{c\Gamma(1 + \tau)} = v^\tau\frac{e^{-(v/c)}}{\Gamma(1 + \tau)c^{\tau+1}}. \qquad (18)$$

Comparing (18) with (17), it can be seen that to equate the two relations the parameters $\tau = \alpha - 1$ and $c = \beta$ can be substituted into (18), yielding

$$v^{\alpha-1}\frac{e^{-(v/\beta)}}{\Gamma(1 + \tau)\beta^\alpha}. \qquad (19)$$

Comparing (19) with (17), it now becomes obvious that with the above new representations for τ and c Eq. (19) becomes a gamma PDF.

In terms of these parameters, the first two central moments of the gamma density function are now $c(\tau + 1)$ and $c^2(\tau + 1)$, respectively. Using data to estimate the first two raw moments,

$$\mu_V(1) = \frac{1}{n}\sum_{i=1}^{n} v(i); \quad \mu_V(2) = \frac{1}{n}\sum_{i=1}^{n} v(i)^2; \qquad (20)$$

values for τ and c can be found from the expressions for the first two central moments (Provost 2005). Since $c(\tau + 1)$ is the first central moment and $c^2(\tau + 1)$ is the second, they can be shown as

$$c = \beta = \frac{\mu_V(2) - \mu_V(1)^2}{\mu_V(1)} \quad \text{and}$$

$$\tau = \alpha - 1 = \frac{\mu_V(1)}{c} - 1. \qquad (21)$$

These values can now be used in expansion (11).

Summarizing, the generalized Laguerre process involves

1) computing the first two raw moments from the data using (20),
2) using (21) to find values for c and τ,
3) computing $K + 1$ expansion coefficients using (16),
4) using (11) to construct $K + 1$ number of expansions, and finally
5) using a suitable fitting assessment routine to determine the optimal number of terms to include in the expansion.

d. The Kronmal–Tarter criterion for degree selection

Kronmal and Tarter (1968), Tarter and Kronmal (1976), and Diggle and Hall (1986) each proposed a method for optimizing the number of terms in the function estimator. They defined a function [i.e., $J(k)$] of the number of terms k to be included in the expansion as an expression of the expected value of the squared difference between the density function estimated with k terms and the true density function. In essence, the kth term should be included in the expansion only if $J(k) < J(k - 1)$. A summary of the technique is given by Jiang and Provost (2011). This study does not use this method, as our objective is to observe the quality of various ‘‘fits’’ of the expansion to data even past the ‘‘optimized’’ number of terms.

3. Experiment

To investigate the utility of the three expansions to produce a tractable, accurate function for the wind speed

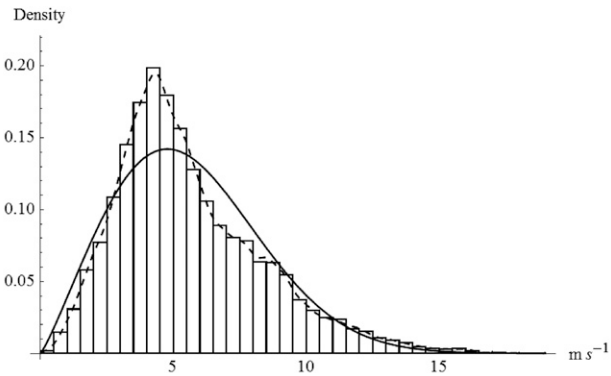

FIG. 1. Histogram of 5-min wind speed recorded at Boise City, OK, for May 2000. A Gaussian kernel (GH) with a bandwidth of 0.4 m s^{-1} is show by the dashed line. The two-parameter Weibull function was fit using the maximum likelihood method and is shown by the solid line.

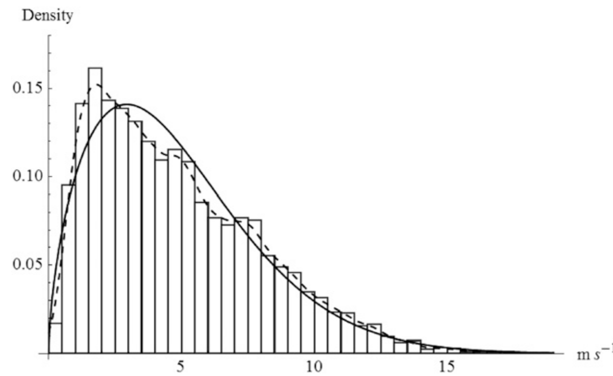

FIG. 2. As in Fig. 1, but at Kenton, OK.

PDF, a sample from two datasets of 5-min wind speed data for May 2000, from Boise City and Kenton were obtained from the Oklahoma Mesonetwork (Brock et al. 1995). Hereinafter, the Laguerre expansion will simply be referred to as the LE expansion, the Gauss–Hermite expansion as the GH expansion, and the generalized Laguerre expansion as the GL expansion. The histograms for Boise City and Kenton are shown in Figs. 1 and 2, respectively. These two sites were selected due to their differences in shape and moments, and their non-Weibull-like frequency histograms. Both sites are located in the far western portion of the Oklahoma panhandle. The specific reasons for their differences in shape are unclear. One possible explanation can be found in terrain differences in the region. Notably there is a 55-m difference in elevation between the two sites, which are separated by approximately 40 km. A Gaussian kernel with a bandwidth of 0.4 m s^{-1} was selected for both for Boise City and Kenton using Silverman's rule (Silverman 1998) and is shown in Figs. 1 and 2 with dashed lines. A maximum likelihood best-fit two-parameter Weibull function is also shown in Figs. 1 and 2 with solid lines. While the Weibull function provides a better fit to the Kenton data than to that of Boise City, it is obviously inferior to the fit using the Gaussian kernel in both cases.

The Gaussian kernel (GK), as expected, proves to be extremely flexible and provides an excellent fit to both datasets. Although not a simple analytic function and thus not useful for the applications listed above, the numerical interpolating function associated with the kernel can be used as a standard function whereby the adequacy of the fit of the three expansions to each dataset can be assessed. To check the fit of the kernel interpolation function itself, it is compared with the Boise City data on a cumulative probability plot in Fig. 3. The dashed one-to-one reference line overlies the probability plot, indicating an extremely good fit throughout the range of data values. The same plot for Kenton (not shown) is nearly identical. Thus, the GK interpolating function for both datasets will be assumed to accurately represent the true PDFs for experimental purposes. The error in the fit of a selected function expansion for a given number of terms using the GK function is determined using MISE:

$$\mathrm{MISE}_T(K+1) = \int_0^\infty [f_T(v, K+1) - \mathrm{GK}(v)]^2\, dv, \qquad (22)$$

where the T in the subscript of $\mathrm{MISE}_T(K + 1)$ and $f_T(v, K)$ refers to the type of expansion (i.e., T = LE, GH, or GL).

a. Boise City

Equation (22) was used to estimate the MISE for each expansion type and number of terms up to $K + 1 = 11$.

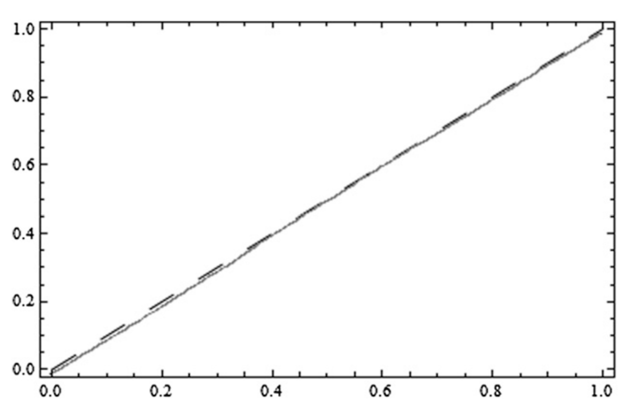

FIG. 3. The cumulative probability plot for the Gaussian kernel interpolating function vs the wind speed data for Boise City. A dashed reference one-to-one line is also plotted.

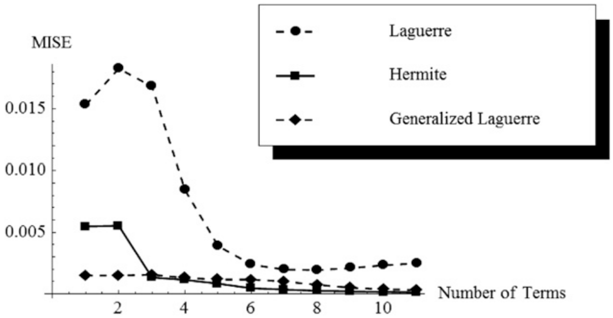

FIG. 4. The MISE (s m^{-1}) computed for each expansion for Boise City as a function of the numbers of terms in the expansion.

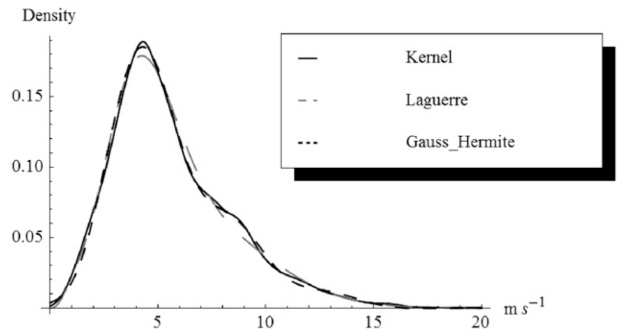

FIG. 5. A comparison of the GH and GL expansions with 10 terms each with the Gaussian kernel interpolation function for Boise City.

As can be observed from Fig. 4, for expansions greater or equal to 3 terms, the GH expansion performs best for Boise City with 10 terms being an acceptable number. While the GL expansion also provides a good fit, it also requires about 10 terms to have a similar MISE as the GH expansion. The LE expansion proves inferior to the two other expansions for all terms included. This is likely due to the relative inflexibility of the basic Laguerre polynomial in comparison with the generalized form, which has an extra parameter. In Fig. 5, the 10-term GH expansion and the 10-term GL expansions are compared with the GK function. Only minor differences are observable between the GH and the GK fits, with the GH expansion providing a closer estimate of the peak density value. The GH expansion with 10 terms takes the following analytical form:

$$f_{\mathrm{GH}}(v) = e^{-0.0662(-5.6231+v)^2}(0.0293 + 0.020v + 0.0219v^2 + 0.007v^3 - 0.005v^4 + 0.0002v^5 + 0.0002v^6 - 0.000\,03v^7 + 0.000\,003v^8 - 1.103\,33\times10^{-7}v^9 + 1.615\,09\times10^{-9}v^{10}). \quad (23)$$

b. *Kenton*

Figures 6 and 7 show the analysis for the distribution fits and MISE comparisons for the different approaches. It can be seen from these figures that both of the Laguerre expansions outperformed the GH expansions throughout the range of terms. Based on MISE values, the GL expansion performed better than the LE expansion after term 3. The optimal number of terms for the GL and LE expansions appears to be about eight with the GL expansion providing the best fit of the three to this dataset. Thus, the GL expansion provides the best fit of the three expansions for this particular dataset. The additional parameter in the generalized Laguerre polynomial helps provide additional flexibility.

The resulting eight-term GL expansion is

$$f_{\mathrm{GL}}(v) = e^{-0.0514(-4.834+v)^2}(0.1256 + 0.1118v - 0.0282v^2 - 0.0066v^3 + 0.0022v^4 - 0.000\,08v^5 - 0.000\,03v^6 + 0.000\,003v^7 - 7.8553\times10^{-8}v^8). \quad (24)$$

4. Gram–Schmidt orthogonalization

In the above section, three different expansions of classical orthogonal polynomials were fit to two datasets with the result that the GL expansion proved superior to the LE expansion in terms of the MISE for these datasets. However, the GH expansion provided a better fit to the more Gaussian-like shaped Boise City histogram. Thus, it is likely that a better fit could be achieved given an entirely different set of orthogonal polynomials. Not all orthogonal polynomials are of the classical types. The Gram–Schmidt process (Arfken and Weber 2005) can produce a set of orthogonal polynomials given any weight function and a set of linearly independent basis functions. For example, given the basis set $\{1, v, v^2, v^3, \ldots\}$ and any square integrable weight function $w(v)$ defined over a given level of support, (a, b), a set of orthogonal polynomials (i.e., $P_k(x); k = 0, 1, 2, \ldots$) can be constructed for any positive, real random variable x, defined over this same level of support using

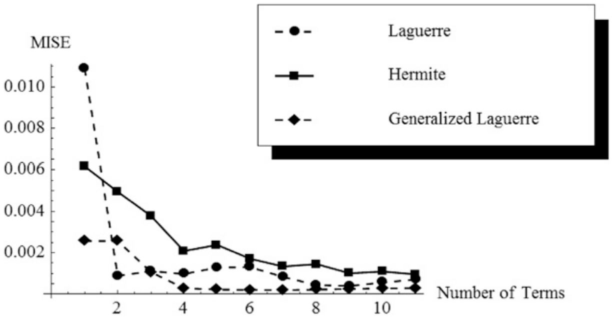

FIG. 6. As in Fig. 4, but for Kenton.

$$P_0(x) = 1$$
$$P_1(x) = x - \frac{\langle x, 1\rangle}{\langle 1, 1\rangle}$$
$$P_{k+1}(x) = \left[x - \frac{\langle xP_k(x)P_L(x)\rangle}{\langle P_k(x)P_k(x)\rangle}\right]P_k(x) - \left[x - \frac{\langle xP_k(x)P_L(x)\rangle}{\langle P_{k-1}(x)P_{k-1}(x)\rangle}\right]P_{k-1}(x), \quad (25)$$

with the notation

$$\langle P_k(x)P_k(x)\rangle = \int_a^b w(x)P_k(x)P_k(x)\,dx. \quad (26)$$

What is more, if the weight function is a proper density function, the polynomials will be orthonormal with respect to that weight function as well.

Using the above process, a set of orthogonal polynomials were constructed using the Weibull density function as the weight function to assess how an expansion of such polynomials compared to the classical types already fit to the two datasets. First, the Weibull function defined for the wind speed random variable V is

$$f_W(v) = \frac{\gamma}{\eta}\left(\frac{v}{\eta}\right)^{\gamma-1} e^{-(v/\eta)^\gamma}; \quad v \in [0, \infty); \gamma, \eta > 0, \quad (27)$$

where γ, η are the shape and scale parameters, respectively. Using the maximum likelihood method, these two parameters turn out to be $\gamma = 2.1615$ and $\eta = 6.3656$ m s^{-1}. Using these values for the parameters in the weight function, $w(v)$, the first few orthogonalized polynomials constructed using (25) over the support $[a = 0, b = \infty)$ are

$$\{1.0, -2.0513 + 0.3639v, 3.2094 - 1.2594v + 0.0989v^2, -4.4366 + 2.8506v - 0.4793v^2 + 0.0227v^3\}. \quad (28)$$

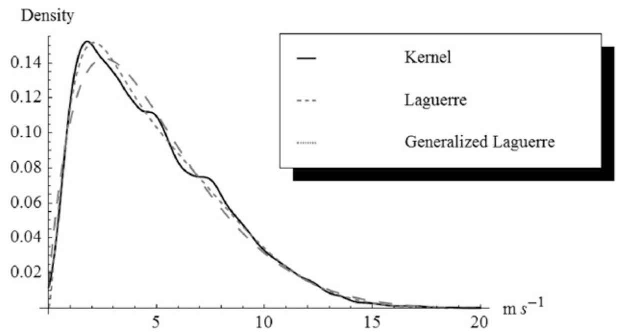

FIG. 7. A comparison of the GL and LE expansions with eight terms each with the Gaussian kernel interpolating function for Kenton.

The resulting expansion takes the form

$$f(v) = w(v)\sum_{k=0}^{K}\psi_k P_k(v), \quad (29)$$

where the coefficients can be estimated from the data using the orthonormal relationship from

$$\psi_k = \int_0^\infty \varphi_k(v)f(v)\,dv \sim \frac{1}{n}\sum_{i=1}^{n} P_k[v(i)]. \quad (30)$$

Note that the data do not have to be transformed as with the classical orthogonal polynomials since the weight function is already a true probability density (i.e., not standardized) whose parameters have been estimated from the data.

A comparison of the fit of the Weibull-weighted (WW) orthogonal expansion with the GH and the GL expansions applied to both datasets is shown in Figs. 8–11. For Boise City, on a term-by-term basis, the GH expansion outperforms the WW expansion. Interestingly, the first term of the WW expansion (i.e., the Weibull PDF) provided a better fit to the data than did the first term of the GH expansion (i.e., the Gaussian PDF). The GL expansion provided the best overall first-term fit. However, after term 2, the additional terms in the GH expansions added significant improvement to the fit and continued to improve the fit through the number of terms included (Fig. 8). Figure 9 shows that the GH expansion estimated the peak of the density quite well.

The MISE results for Kenton are shown in Fig. 10. Overall, the GL expansion performed better than the WW expansion after the second term. Both expansions were superior to the GH expansion for Kenton. Figure 10 also shows a convergence in the MISE with increasing number of terms for the WW and GL expansions. Figure 11 indicates that the GL captures the peak density slightly better than did the WW expansion.

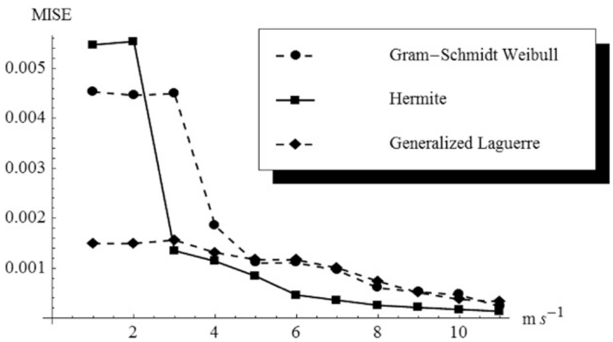

FIG. 8. The MISE (s m^{-1}) computed for each expansion for Boise City as a function of the numbers of terms in the expansion. Included is the expansion developed from the Gram–Schmidt process using the Weibull function as the basis weight function.

5. Conclusions

The objective of this study was to determine if relatively simple analytic expressions could be used to approximate somewhat ''ill behaved'' wind speed PDFs, and also provide a more robust and systematic estimator than the more traditional Weibull function fit. By ill-behaved it is meant those PDFs whereby the two-parameter Weibull function provides a poor fit. It was noted that while a kernel interpolation function is extremely flexible and, thus, is useful for many research applications, there are situations where a simple analytic expression is required, particularly as applied to downscaling climate models or wind turbine engineering applications. Morrissey et al. (2010a) showed how a Gauss–Hermite expansion of the wind speed PDF can be expressed as a function of a scaling parameter. The scaling parameter itself is a function of the second moment of the wind speed and thereby can be used to adjust the ''shape'' of the wind speed PDF since wind speed variance is directly related to measurement resolution. Thus, near-optimal orthogonal expansions of the wind speed

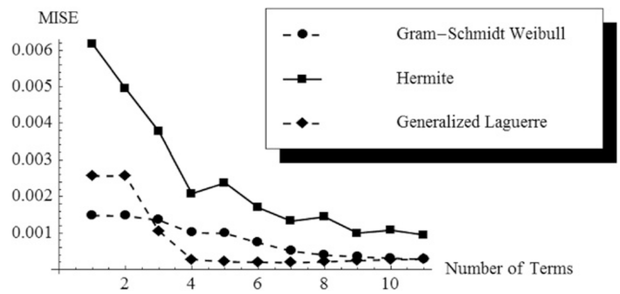

FIG. 10. As in Fig. 8, but for Kenton.

PDFs will have direct and useful relevance to downscaling and similar research topics.

Three different expansions of two classical orthogonal polynomials were given together with an expansion of orthogonal polynomials constructed using the Gram–Schmidt method with the Weibull PDF as the weight function. By applying the GK interpolating function to two different near-surface (10 m) wind speed datasets, it was found that the GK interpolating function fits are good enough to use as a ''standard'' against which to compare the four analytical expansions. It was demonstrated that the GL expansion proved superior to the LE on a term-by-term basis for the datasets. This was likely due to the extra parameter in the generalized Laguerre polynomial. Interestingly, the GH expansion outperformed the GL expansion for the Boise City data. This supports the conjecture by Provost (2005) that the basic underlying shape of the wind speed histogram is important in determining which expansion proves superior. The Boise City wind speed histogram is more ''Gaussian'' in shape than the histogram for Kenton, which has a more gammalike shape.

From the experiments given in this study it has been shown that the closer a weight function is to providing a first ''best guess'' initial fit to the histogram, the better the fit of the associated expansion. It is interesting that

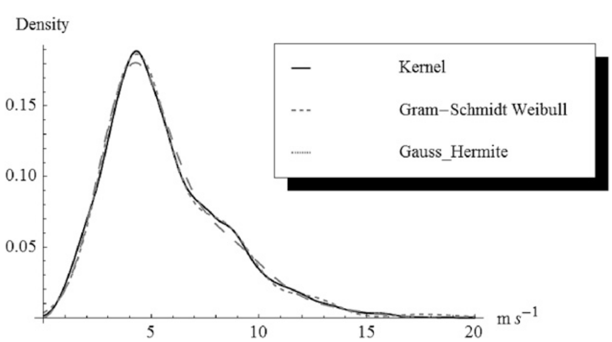

FIG. 9. A comparison of the Gram–Schmidt Weibull expansion with the GL expansion with 11 terms each with the Gaussian kernel interpolating function for Boise City.

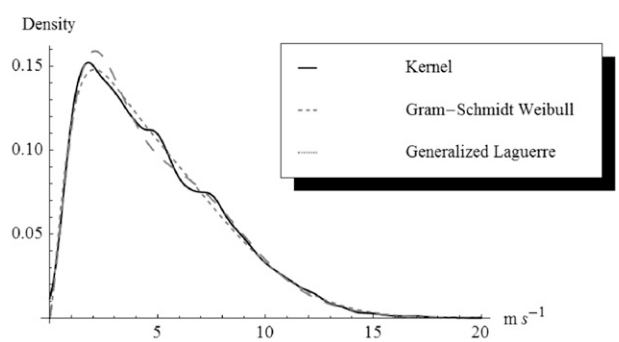

FIG. 11. A comparison of the Gram–Schmidt Weibull expansion with the GL expansion with nine and six terms, respectively, with the Gaussian kernel interpolating function for Kenton.

given a Weibull PDF as a weight function, additional terms in the expansion generally improved the fit. Thus, since additional terms in an expansion act to adjust the fit and the first term in an expansion is the weight function, then there will always be an orthogonal polynomial expansion that either meets or exceeds the accuracy of fit of a Weibull function PDF. In summary, if a tractable analytic function for the wind speed PDF is required, then it is likely that one of the above classical orthogonal expansions will prove useful. One can always experiment with the Gram–Schmidt orthogonalization method to construct expansions that may provide useful fits to otherwise very unusually shaped wind speed histograms.

Acknowledgments. The authors acknowledge the support of the U.S. Department of Energy, the Oklahoma Department of Commerce, and the Oklahoma Economic Development Generating Excellence program in providing financial support for this research. The authors also thank Professor Serge Provost at the University of Western Ontario for his helpful and insightful comments on the derivations and usage of the orthogonal polynomial expansions used in this work.

REFERENCES

Arfken, G. B., and H. J. Weber, 2005: *Mathematical Methods for Physicists, Sixth Edition: A Comprehensive Guide*. Academic Press, 1200 pp.

Baker, R. W., and J. P. Hennessey Jr., 1977: Estimating wind power potential. *Power Eng.,* **81,** 56–57.

Barthelmie, R. J., and S. C. Pryor, 2003: Can satellite sampling of offshore wind speeds realistically represent wind speed distributions? *J. Appl. Meteor.,* **42,** 83–94.

Blinnikov, S., and R. Moessner, 1998: Expansion for nearly Gaussian distributions. *Astron. Astrophys. Suppl. Ser.,* **130,** 193–205.

Brock, F. V., K. C. Crawford, R. L. Elliott, G. W. Cuperus, S. J. Stadler, H. L. Johnson, and M. D. Eilts, 1995: The Oklahoma Mesonet: A technical overview. *J. Atmos. Oceanic Technol.,* **12,** 5–19.

Bryukhan, F. F., and R. D. Diab, 1993: Decomposition of empirical wind speed distributions by Laguerre polynomials. *Wind Eng.,* **17,** 147–151.

Cakmur, R. V., and R. L. Miller, 2004: Incorporating the effect of small-scale circulations upon dust emission in an atmospheric general circulation model. *J. Geophys. Res.,* **109,** D07201, doi:10.1029/2003JD004067.

Capps, S. B., and C. S. Zender, 2008: Observed and CAM3 GCM sea surface wind speed distributions: Characterization, comparison, and bias reduction. *J. Climate,* **21,** 6569–6585.

Diggle, P. J., and P. Hall, 1986: The selection of terms in an orthogonal series density estimator. *J. Amer. Stat. Assoc.,* **81,** 230–233.

Duenas-Osorio, L., and B. Basu, 2008: Unavailability of wind turbines due to wind-induced accelerations. *Eng. Struct.,* **30,** 885–893.

Feely, R. A., R. Wanninkhof, W. McGillis, M.-E. Carr, and C. E. Cosca, 2004: Effects of wind speed and gas exchange parameterizations on air–sea CO_2 fluxes in the equatorial Pacific Ocean. *J. Geophys. Res.,* **109,** C08S03, doi:10.1029/2003JC001896.

Frech, M., F. Holzapfel, A. Tafferner, and T. Gerz, 2007: High-resolution weather database for the terminal area of Frankfurt airport. *J. Appl. Meteor. Climatol.,* **46,** 1913–1932.

Gustafson, W. I., Jr., and L. R. Leung, 2007: Regional downscaling for air quality assessment: A reasonable proposition? *Bull. Amer. Meteor. Soc.,* **88,** 1215–1227.

Hennessey, J. O., 1977: Some aspects of wind power statistics. *J. Appl. Meteor.,* **16,** 119–128.

Jaramillo, O. A., and M. A. Borja, 2004: Wind speed analysis in La Ventosa, Mexico: A bimodal probability distribution case. *Renew. Energy,* **29,** 1613–1630.

Jiang, M., and S. Provost, 2011: Improved orthogonal polynomial density estimates. *J. Statist. Comput. Simul.,* **81,** 1495–1516.

Johnson, N., S. Kotz, and N. Balakrishnan, 1994: *Continuous Univariate Distributions*. Vol. 1, 2nd ed., Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, 756 pp.

Juban J., N. Siebert, and G. N. Kariniotakis, 2007: Probabilistic short-term wind power forecasting for the optimal management of wind generation. *Proc. PowerTech 2007,* Lausanne, Switzerland, IEEE, 683–688.

Justus, C. G., W. R. Hargraves, A. Mikhail, and D. Graber, 1978: Methods for estimating wind speed frequency distributions. *J. Appl. Meteor.,* **17,** 350–353.

Kronmal, R., and M. Tarter, 1968: The estimation of probability densities and cumulatives by Fourier series methods. *J. Amer. Stat. Assoc.,* **63,** 925–952.

Lackner, M. A., A. L. Rogers, and J. F. Manwell, 2008: Uncertainty analysis in MCP-based wind resource assessment and energy production estimation. *J. Sol. Energy Eng.,* **130,** 031006, doi:10.1115/1.2931499.

Li, M., and X. Li, 2005: MEP-type distribution function: A better alternative to Weibull function for wind speed distributions. *Renew. Energy,* **30,** 1221–1240.

Mearns, L. O., I. Bogardi, F. Giorgi, I. Matyasovszky, and M. Palecki, 1999: Comparison of climate change scenarios generated from regional climate model experiments and statistical downscaling. *J. Geophys. Res.,* **104,** 6603–6621.

Monahan, A. H., 2006a: The probability distribution of sea surface wind speeds. Part I: Theory and sea winds observations. *J. Climate,* **19,** 497–520.

——, 2006b: The probability distribution of sea surface wind speeds. Part II: Dataset intercomparison and seasonal variability. *J. Climate,* **19,** 521–534.

——, 2007: Empirical models of the probability distribution of sea surface wind speeds. *J. Climate,* **20,** 5798–5814.

——, Y. He, N. McFarlane, and A. Dai, 2011: The probability distribution of land surface wind speeds. *J. Climate,* **24,** 3892–3909.

Morrissey, M. L., A. Albers, J. S. Greene, and S. Postawko, 2010a: An isofactorial change-of-scale model for the wind speed probability density function. *J. Atmos. Oceanic Technol.,* **27,** 257–273.

——, W. E. Cook, and J. S. Greene, 2010b: An improved method for estimating the wind power density distribution function. *J. Atmos. Oceanic Technol.,* **27,** 1153–1164.

Pavia, E. G., and J. J. O'Brien, 1986: Weibull statistics of wind speed over the ocean. *J. Climate Appl. Meteor.,* **25,** 1324–1332.

Provost, S. B., 2005: Moment-based density approximants. *Math. J.,* **9,** 727–756.

——, and M. Jiang, 2012: Orthogonal polynomial density estimates: Alternative representation and degree selection. *Int. J. Comput. Math. Sci.,* **6,** 17–24.

Pryor, S. C., M. Nielsen, R. J. Barthelmie, and J. Mann, 2004: Can satellite sampling of offshore wind speeds realistically represent wind speed distributions? Part II: Quantifying uncertainties associated with distribution fitting methods. *J. Appl. Meteor.,* **43,** 739–750.

——, J. T. Schoof, R. J. Barthelmie, 2005: Empirical downscaling of wind speed probability distributions. *J. Geophys. Res.,* **110,** D19110, doi:10.1029/2005JD005899.

Ramirez, P., and J. A. Carta, 2005: Influence of the data sampling interval in the estimation of the parameters of the Weibull wind speed probability density distribution: A case study. *Energy Convers. Manage.,* **46,** 2419–2438.

Silverman, B. W., 1998: *Density Estimation*. Chapman and Hall, 175 pp.

Szego, G., 2003: *Orthogonal Polynomials*. 4th ed. American Mathematical Society Colloquium Publications, Vol. 23, 432 pp.

Tarter, M. E., and R. A. Kronmal, 1976: An introduction to the implementation and theory of nonparametric density estimation. *Amer. Stat.,* **30,** 105–112.

Veers, P. S., 1983: A general method for fatigue analysis of vertical axis wind turbine blades. Sandia National Laboratories Tech. Rep. SAND82-2543, 12 pp.

Vickers, D., and S. Esbensen, 1998: Subgrid surface fluxes in fair weather conditions during TOGA COARE: Observational estimates and parameterizations. *Mon. Wea. Rev.,* **126,** 620–633.

Wang, C., R. H. Weisberg, and H. Yang, 1998: Effects of the wind speed–evaporation–SST feedback on the El Niño–Southern Oscillation. *J. Atmos. Sci.,* **56,** 1391–1403.

White, D., 2004: New method for dual-axis fatigue testing of large wind turbine blades using resonance excitation and spectral loading. National Renewable Energy Laboratory Tech. Rep. NREL/TP-500-35268, Golden, CO, 195 pp.

Wright, D. G., and K. R. Thompson, 1983: Time-averaged forms of the nonlinear stress law. *J. Phys. Oceanogr.,* **13,** 341–345.

Yilmaz, V., and H. E. Çelik, 2008: A statistical approach to estimate the wind speed distribution: The case of Gelibolu region. *Doğuş Üniv. Derg.,* **9,** 122–132.

Zagar, N., M. Zagar, J. Cedilnik, G. Gregoric, and J. Rakovec, 2006: Validation of mesoscale low-level winds obtained by dynamical downscaling of ERA40 over complex terrain. *Tellus,* **58A,** 445–455.