Environmental Research Letters

OPEN ACCESS

RECEIVED
9 December 2014

REVISED
12 March 2015

ACCEPTED FOR PUBLICATION
13 March 2015

PUBLISHED
2 April 2015

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

LETTER

Variability of interconnected wind plants: correlation length and its dependence on variability time scale

Clara M St. Martin[1], **Julie K Lundquist**[1,2] **and Mark A Handschy**[3]

[1] Department of Atmospheric and Oceanic Sciences, University of Colorado at Boulder, Boulder, CO, USA
[2] National Renewable Energy Laboratory, Golden, CO, USA
[3] Enduring Energy, LLC, Boulder, CO, USA

E-mail: clara.st.martin@colorado.edu

Keywords: wind power, variability, geographic diversity

Supplementary material for this article is available online

Abstract

The variability in wind-generated electricity complicates the integration of this electricity into the electrical grid. This challenge steepens as the percentage of renewably-generated electricity on the grid grows, but variability can be reduced by exploiting geographic diversity: correlations between wind farms decrease as the separation between wind farms increases. But how far is far enough to reduce variability? Grid management requires balancing production on various timescales, and so consideration of correlations reflective of those timescales can guide the appropriate spatial scales of geographic diversity grid integration. To answer 'how far is far enough,' we investigate the universal behavior of geographic diversity by exploring wind-speed correlations using three extensive datasets spanning continents, durations and time resolution. First, one year of five-minute wind power generation data from 29 wind farms span 1270 km across Southeastern Australia (Australian Energy Market Operator). Second, 45 years of hourly 10 m wind-speeds from 117 stations span 5000 km across Canada (National Climate Data Archive of Environment Canada). Finally, four years of five-minute wind-speeds from 14 meteorological towers span 350 km of the Northwestern US (Bonneville Power Administration). After removing diurnal cycles and seasonal trends from all datasets, we investigate dependence of correlation length on time scale by digitally high-pass filtering the data on 0.25–2000 h timescales and calculating correlations between sites for each high-pass filter cut-off. Correlations fall to zero with increasing station separation distance, but the characteristic correlation length varies with the high-pass filter applied: the higher the cut-off frequency, the smaller the station separation required to achieve de-correlation. Remarkable similarities between these three datasets reveal behavior that, if universal, could be particularly useful for grid management. For high-pass filter time constants shorter than about $\tau = 38$ h, all datasets exhibit a correlation length ξ that falls at least as fast as τ^{-1}. Since the inter-site separation needed for statistical independence falls for shorter time scales, higher-rate fluctuations can be effectively smoothed by aggregating wind plants over areas smaller than otherwise estimated.

1. Introduction

Low CO_2 emission footprints make wind and solar power attractive choices for future electricity needs. However, their natural variability is challenging for the electric grid, which requires instantaneous matching of generation and load on all time scales from one ac cycle, through operational scheduling horizons (a day or two), out to planning horizons of more than a decade (von Meier 2006). Variability can be reduced and the grid-integration challenge lessened by interconnecting renewable electricity generators distanced enough that their variation is not fully correlated (Thomas 1945, Kahn 1979). At a great enough distance, which depends on variability time scale as we see here, the wind-speed variations approach statistical independence: an ensemble of wind plants separated at least this distance from each other is 'geographically

© 2015 IOP Publishing Ltd

diverse,' and at the associated time scale, variability of the ensemble's summed power output will be reduced compared to that of an individual plant.

Better plans for future grids could be made if this effect were better understood. Many previous studies have relied on an empirical approach where the degree of smoothing that might have occurred over a specific region was estimated from historical meteorological or reanalysis data or even calculated from actual renewable generation (Molly 1977, Palutikof *et al* 1990, Landberg 1997, Milligan and Factor 2000, Wiemken *et al* 2001, Archer and Jacobson 2003, 2007, Holttinen 2005, Sinden 2007, Kiss and Jánosi 2008, Kempton *et al* 2010, Fertig *et al* 2012, Cosseron *et al* 2013, Fisher *et al* 2013, Huang *et al* 2014, Louie 2014). However, it is often difficult to generalize the findings from the study scenario to realistic planning scenarios. As a first step, some have sought to model how correlation falls off with distance, using a variety of different forms: exponential, both with (Haslett and Raftery 1989, Landberg 1999, Kiss and Jánosi 2008, and Katzenstein *et al* 2010) and without (Giebel 2000, Simonsen and Stevens 2004, Holttinen 2005, Achberger *et al* 2006, and Kempton *et al* 2010) a nugget (non-unity correlation at zero separation), stretched exponentials (Hasche 2010, Louie 2014), Gaussian forms (Buell 1972) and Cauchy forms (Buell 1972, Julian and Thiebaux 1975); see table 1. A few have gone further to propose forms predicting probability distributions of aggregated power as a function of the region size (Justus and Mikhail 1978, Carlin and Haslett 1982, Haslett and Raftery 1989, Hasche 2010).

How correlations depend on the variability time scale is equally important to how they depend on distance, since, for a given magnitude operational power excursion, the faster the excursion generally the more costly its regulation (Kirby 2004). Without suggesting a particular functional form for the dependence, studies by Ernst *et al* (1999) and by Mills and Wiser (2011) have found for wind and solar, respectively, that faster variations become uncorrelated at smaller spatial separations than slower variations. Variability time scale τ or frequency f can be inserted into functional forms for correlation ρ versus distance r by introducing a characteristic velocity v, and replacing correlation length ℓ by $v\tau$ or v/f, as proposed by Davenport (1961). Beyer *et al* (1990, 1993), McNerney and Richardson (1992), Nanahara *et al* (2004), and Sørensen *et al* (2008) have taken this approach for wind-speed correlation. Calling $r/(v\tau)$ a 'dispersion factor', Hoff and Perez (2010) introduced this concept to the study of solar variability and identified velocity v with cloud motion; it has since been used in solar correlation studies by Marcos *et al* (2012), Lave and Kleissl (2013), and Hinkelman (2013). None of the studies based on characteristic-velocity functional forms have considered regions much larger than 100 km in extent, and more typically have been limited to the size of a single PV plant.

Here we revisit the methods used by Ernst *et al* (1999) and by Mills and Wiser (2011), apply them to both wind-speed data and wind generation data from three different geographical regions, and find a single quantitative relationship between correlation length and time scale that parsimoniously characterizes behavior on time scales ranging from an hour to a quarter-year and over distances of a kilometer to a continent.

In section 2, we highlight the unique features of the datasets used in this analysis as well as novel methods for investigating correlations between sites and quantifying correlation length scales. Section 3 describes the results found in our investigation and section 4 sums up our conclusions and suggestions for future work. In the provided supplementary data (stacks.iop.org/ERL/10/044004/mmedia), we describe, in detail, the filtering methodology used to eliminate non-stationarities in the time series and present the correlation results from all datasets.

2. Data and methods

2.1. Datasets

Although previous studies have characterized correlations within an individual region, the present study seeks universal behavior across multiple regions and datasets: an Australian wind generation dataset (AUS), a Canadian wind speed dataset (CAN) and a Bonneville Power Administration wind speed dataset (BPA). The three datasets presented here bring different features to the study, such as wind power production data (AUS), great extent (CAN) or fine time resolution (AUS, BPA). More details on these datasets, including maps of the stations used, appear in the supplementary data (stacks.iop.org/ERL/10/044004/mmedia).

2.1.1. Australian generation dataset (AUS)

The Australian Energy Market Operator (AEMO) provides a 1 year (October 2013–October 2014) dataset of five-minute electricity generation data from 32 wind farms across Southeastern Australia including the provinces of New South Wales, South Australia, Tasmania and Victoria (table 2). AEMO reported in 2013 that they might curtail wind power systems in the future (Australian Energy Market Operator 2013); therefore we suspect few curtailment effects in this dataset. We utilize a subset of 29 wind farms, combining production for different wind farm stages and selecting only the larger-producing member of wind farm pairs within 5 km of one another. We make no correction for wake effects of neighboring wind farms (Fitch *et al* 2013) in this analysis.

The terrain elevations of these wind farms range from 3 to 900 m above sea level (asl). Plant output data are normalized by each wind plant's generation capacity, which range from 20 to 420 MW. The closest pair of farms is separated by 8 km and the farthest by 1274 km.

Table 1. Correlation functions found in previous work.

Source	Functional form (r is distance)	Adjustable parameters	Length (km)	Data description
Buell (1972)	Longitudinal: $\exp\left[-\frac{1}{2}(\alpha r)^2\right]$	α		500 mb wind data from Europe and the North Atlantic during the summer
	Transverse: $[1+(\alpha r)^2]^{-q-1}$,	α, q		
	$\left[1-(\alpha r)^2\right]\exp\left[-\frac{1}{2}(\alpha r)^2\right]$,	α		
	$\left[1-(2q+1)(\alpha r)^2\right]\left(1+(\alpha r)^2\right)^{-q-2}$	α, q		
Julian and Thiebaux (1975)	$[\alpha\cos(\omega r)+z-\alpha]\left[1+(\lambda r)^2\right]^{-\gamma}$	$\alpha, \omega, \lambda, \gamma, z$		2 yr of 500 mb wind data from North America during the winter (3 months)
Justus and Mikhail (1978)	$\alpha\exp(-r/L)$[a]	α, L	$\alpha L = 395$–560[a]	(fit to data in J&M figure B-4) 1 yr of 3 hrly data at 10 m from 22 US sites
Haslett and Raftery (1989)	$\alpha\exp(-\beta r)$	α, β	$\alpha\beta^{-1} = 722$	18 yr of hrly mean wind speeds from 12 Ireland stations
Beyer *et al* (1993)	$\exp\left[-(kr)^m\right]$	k, m	$k^{-1} = 870$	3 yr of hrly mean wind speeds (converted to power) from four sites in NW Germany
Landberg (1999)	$\exp(-r/a)-\beta$	a, β	$(1-\beta)a = 425$	1 yr of 10 min power output from 17 wind farms in Denmark
Giebel (2000)	$\exp(-r/D)$	D	$D = 723$	1 yr of hrly 50 m wind speeds from 60 European stations
Simonsen and Stevens (2004)	$\exp(-\alpha r)$	α	$\alpha^{-1} = 323$–588	1 yr of hrly averaged 50 m wind speeds (converted to power) from 28 Midwest US sites
Holttinen (2005)	$\exp(-r/a)$	a	$a = 500$	1 yr of hrly wind power output from 100s of Nordic sites
Gibescu *et al* (2006)	$\alpha\exp(-\beta r)$	a, β		1 yr of 10 min averaged wind speeds for 18 sites in the Netherlands
Achberger *et al* (2006)	$\exp\left(-r/x_0\right)$	x_0	$x_0 = 310$[a]	(curve in Achberger figure 6(a)) 2 yr of 3 hrly 10 m winds From 142 sites in Sweden
Kiss and Jánosi (2008)	$\exp(-r/L)$	L	$L = 400$–600	44 yr of 6 hrly wind (converted to power) from ECMWF ERA-40 model output
Adams and Cadieux (2009)	$\exp(-\alpha r)$	α	$\alpha^{-1} = 333$	2 yr of hrly production from wind farms in Ontario
Hasche (2010)	$\exp\left(-ar^{\alpha}\right)$	a, α	$a^{-(1/\alpha)} = 322$;	1 yr of hrly 10 m wind speeds (converted to power) from 24 German sites
			$a^{-(1/\alpha)} = 22$	As above, 2 h ramps
Katzenstein *et al* (2010)	$\alpha\exp(-r/D)$	α, D	$\alpha D = 271$	15 min power output from 20 wind farms in Texas
Kempton *et al* (2010)	$\exp(-r/D)$	D	$D = 430$	5 yr of (5–49 m) wind data (converted to power) from 11 stations on the US East Coast
Baïle and Muzy (2010)	$\lambda^2\ \ln(L/r)$	λ, L	$L = 600$	17 yr of hrly 10 m wind at 27 locations in the Netherlands
Šaltytė Benth and Šaltytė (2011)	$\exp\left(\sqrt{\theta_2 x^2+y^2}/\theta_1\right)$	θ_1, θ_2	$\theta_1 = 370$	($r^2 = x^2+y^2$) 31 yr of daily wind data from 18 stations in Lithuania
Hill *et al* (2012)	$\alpha\exp(-r/L)$	α, L	$\alpha L = 340$	Multiple y of wind speed from 14 sites in the UK
Louie (2014)	$\exp\left(-\alpha r^{\beta}\right)$	α, β	$\alpha^{-(1/\beta)} = 1100$	3 yr of hrly power output from US system operators

[a] Our fit to cited data.

Table 2. Description of datasets used in this work.

Dataset	AUS	CAN	BPA
Location	SE Australia	Canada	NW US
Type of data	Wind farm generation	Wind speed	Wind speed
Height	Hub height	10 m	6–53 m
Time resolution	5 min	1 h	5 min
Duration	1 yr	45 yr	4 yr
Horizontal extent	1274 km	5344 km	354 km
Closest Site Pair	8 km	19 km	13 km
Number of stations	29	117	14

2.1.2. Canadian wind speed dataset (CAN)

The National Climate Data Archive of Environment Canada provides a 54 year dataset of hourly surface wind-speeds from 117 stations ranging across more than 5000 km of Canada (table 2). While the entire dataset spans 1953 through 2006, missing data are less frequent in later years, so we use a 45 year subset of data, from 1962 to 2006. During the data collection, the measurement heights of the stations ranged from 5 to 32 m above ground level (agl), but Wan *et al* (2010) homogenized these data to the standard 10 m height assuming a logarithmic wind profile. The terrain elevations within this dataset vary from sea level to 1084 m asl. Separation distances between sites vary from a minimum of 19 km to a maximum of 5344 km.

2.1.3. BPA wind speed dataset

The BPA provides a dataset of five-minute wind-speeds from meteorological towers in Washington and Oregon in the Northwestern US (table 2). For this analysis, we use data from 14 sites from 2010 to 2014 due to the higher temporal resolution of these data, although several stations provide data for longer durations with lower time resolution (BPA).

These 14 towers in the BPA network are generally located along the Columbia River Gorge on the border of Washington and Oregon or along the Pacific Coast. The elevations of these stations range from 19 to 1261 m asl. The anemometer heights range from 9 to 53 m agl. No 'standard height' homogenization methodology is used on this dataset; correlation coefficient calculations between sites would be unchanged if either a logarithmic wind profile was assumed or the wind power law was used to homogenize the measurements to a 10 m height. Separation distances between sites vary from a minimum of 13 km to a maximum of 354 km.

2.2. Data processing

2.2.1. Pre-processing

Although both deterministic variability, such as diurnal or seasonal cycles, and random or stochastic variability will figure in any complete accounting of wind power variability, it is the stochastic phenomena that are less understood and not as predictable, so we focus on them here. The analysis of underlying stochastic components is complicated by the presence of temporal periodicities, as noted by Haslett and Raftery (1989), Gunst (1995), Robeson and Shein (1997), Achberger *et al* (2006), and Hill *et al* (2012). To remove the diurnal cycle in order to more clearly characterize stochastic spatial correlations, we make a 'local' estimate of the amplitudes of the first four daily harmonics by least-squares fitting using a 90 day moving window to allow for seasonal variation in the daily cycle (Baïle *et al* 2011). This cycle, estimated once for each of the days in the dataset, is subtracted from that day's wind-speed or power data, as further described in the supplementary data. By also subtracting the 90 day moving average, we remove low-frequency variability, including the seasonal or annual cycle.

As shown by previous investigators (Ernst *et al* 1999), rapid wind-speed or power variations become uncorrelated at smaller spatial separations than do slow variations. In order to further investigate this effect, we prepare versions of each time series with trends slower than a chosen time-constant removed by calculating the average of the wind speed or power data over a segment centered at each time point and subtracting the segment average from the value at the center point. By varying the width of the segment ('window width'), we control the corner frequency of this high-pass filtering operation. Our process, described further in the supplementary data (stacks.iop.org/ERL/10/044004/mmedia), differs from that used by Ernst *et al* (1999), Mills and Wiser (2011), and Hinkelman (2013) in that it subtracts from the original data an averaged version of that data, rather than calculating 'ramps' by differencing one block average value from the next. Thus, the number of data points in the filtered versions of our time series is reduced from the original only by the width of filter window, rather than being decreased by a factor of the window width. We analyze the effects of high-filtering for window widths $\tau \leqslant 1049$ h, and denote by $\tau = 2160$ h data from which the 90 day seasonal bias term has been subtracted but no further filtering applied.

2.2.2. Correlation length

After pre-processing the data as described above and discussed in more detail in the supplementary data (stacks.iop.org/ERL/10/044004/mmedia), we calculate correlations between each pair of stations for each high-pass filtered version of the data. The scatter plot in figure 1, for example, shows the correlation data for CAN at filter window $\tau = 65$ h. Given the high degree of scatter shown here, estimating a single correlation length for each dataset at each filtering window τ is problematic since a functional form of correlation versus distance appropriate for fitting is not known *a priori*.

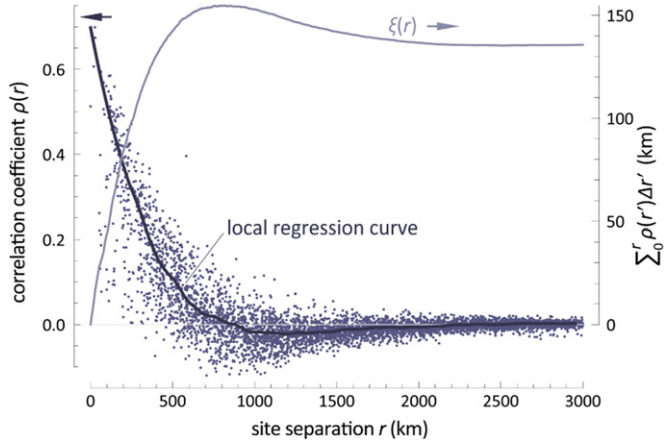

Figure 1. Correlation versus distance for CAN data high-pass filtered with $\tau = 65$ h (scatter plot and dark-blue local-regression curve, left axis). Numerically integrated correlations as a function of integration range (integration of scatter points, light-blue curve, right axis).

Wishing to avoid the functional-form conundrum, we sought a 'non-parametric' distance measure. For a given averaging window width τ we order the correlation coefficient values ρ_{ij} by increasing station separation distance, with increasing values of separation distance signified with single index as r_k, r_1 denoting the smallest separation (the closest pair of stations) and r_N denoting the largest separation (the farthest pair). We then numerically 'integrate' the correlation data over distance using the trapezoid rule:

$$\xi(r_n) = \sum_{k=1}^{n} \frac{1}{2}(r_k - r_{k-1})\left(\rho_{r_k} + \rho_{r_{k-1}}\right). \qquad (1)$$

In the familiar case $\rho(r_i) = \exp(-r_i/\ell) + \varepsilon_i$, where ε_i is mean-zero noise, this procedure gives $\xi = \ell$ in the limit of large r. For our observations, we take $r_0 \equiv 0$, and estimate ρ_{r_0} as the intercept of a local regression curve (see section 2.2.3 below). As shown by the light blue line in figure 1, $\xi(r_n)$ varies smoothly with r_n and approaches a near-constant value at the largest separations (here 139 km at $r_N = 5300$ km) as the ρ_{r_k} scatter there around zero. We refer to the value at the greatest separation $\xi(r_N)$ (at a given high-pass filter window width τ hereinafter ξ_τ) as the *correlation length*; it has a meaning similar to Bretherton's 'correlation radius' (1999) or the term 'integral scale' in turbulence (Batchelor 1953). Our methodology bears some resemblance to that of Şen's cumulative semivariogram (Şen and Şahin 1997, Şen 1989).

2.2.3. Local regression curves

To facilitate graphical comparisons of correlation versus distance behaviors that might otherwise be obscured by the high degree of scatter in the individual $\rho(r_k)$ values, we also calculate $\rho(r)$ curves using local regression techniques (Cleveland 1979), where, to find the curve value at a given site-separation distance r', we fit a 2nd-order polynomial to the fraction α of the correlation data points that are closest to the given distance (for example in the CAN dataset and choosing $\alpha = 0.05$, the 339 correlation data points out of 6786 total having the smallest values of $|r - r'|$). This gives results as exemplified by the thick blue line in figure 1.

2.2.4. Importance of pre-processing

The high-pass filtering and diurnal-cycle removal (outlined in detail in the supplementary data (stacks.iop.org/ERL/10/044004/mmedia) have significant effects on the correlation behavior, as shown in figure 2, which, to facilitate comparison, portrays correlation versus distance as local regression curves. With no pre-processing (raw data, dashed brown curve), correlation doesn't fall below about 0.04 at the largest site separations, while with the diurnal cycle and seasonal bias removed (solid brown curve), correlation falls to zero. For raw data high-pass-filtered with $\tau = 33$ h (aqua), the correlation does not fall below about 0.10 (dashed curve) for separations smaller than 4000 km, while if the high-pass filter is applied to data from which the diurnal cycle has first been removed, correlation falls to near zero for separations of 600 km (solid curve). High-pass cutoffs of $\tau = 25$–37 h give similarly high correlation floors on raw data while the height of the floor decreases noticeably for cutoffs smaller than 21 h or larger than 41 h.

3. Results and discussion

3.1. Correlations

Despite the differences in methodology, the behavior of correlation with distance and high-pass filtering window width qualitatively resembles the previous wind-power results of Ernst *et al* (1999) and insolation work of Mills and Wiser (2011): correlation falls off more rapidly with site separation the smaller the window-width τ, as seen for the CAN data in figure 3. Interestingly, we observe that for some high-pass filter window widths (for example, a 65 h window width in CAN), the correlations actually become negative

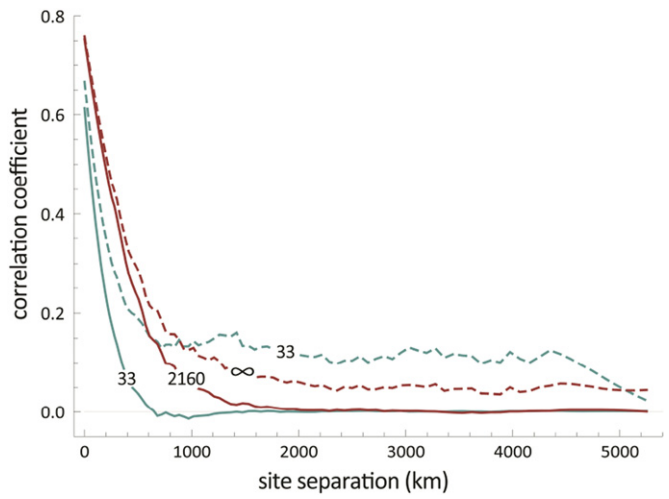

Figure 2. Effect of diurnal-cycle and low-frequency removal on 1962–2006 CAN hourly data correlation (local regression curves). Brown: no pre-processing (dashed, ∞); diurnal-cycle and seasonal bias removed (solid, 2160). Aqua: $\tau = 33$ h high-pass filtered with diurnal cycle subtracted (solid, 33) and without diurnal-cycle subtracted (dashed, 33).

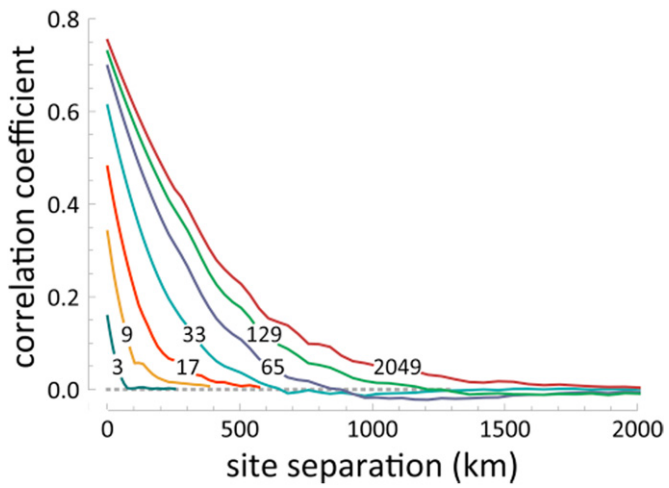

Figure 3. CAN correlation coefficients versus site separation simplified by robust local regression for multiple indicated high-pass window widths in hours.

between 500 and 1000 km separation distances, perhaps for reasons similar to those identified for solar correlations by Hoff and Perez (2010) and Hinkelman (2013).

Substantial scatter of the correlation versus distance values is a prominent feature of data from all three regions at all but the shortest filter widths (see supplementary figures S5–7). We explored several potential causes of this large scatter as well as the negative correlations by separating stations by region (North versus South as well as East versus West) and by azimuthal bearing. East/West differences (figure 4(a)) and North/South differences (figure 4(b)) explain neither the scatter nor the slight anti-correlations. It is clear from previous work (Buell 1972, Ramanathan *et al* 1973, Julian and Thiebaux 1975) that the longitudinal and transverse horizontal wind components have different correlation behaviors, with the transverse correlation falling faster and exhibiting more negative values, similar to the behavior of fully developed turbulence (von Kármán 1948, Batchelor 1953). Unfortunately, the CAN dataset provides wind speed but not wind direction. Nevertheless, some of the scatter in the CAN correlations can be attributed to the azimuthal bearing of each station pair, as stations separated along a line 10° North of West–East are systematically less correlated and more often anti-correlated than those perpendicular to that line (figure 4(c)). Šaltytė Benth and Šaltytė (2011) also observed directional anisotropy in wind-speed correlation fall off. We presume the anisotropy we observe arises from the prevailing westerly winds produced by the large-scale circulation in this region, although, contrary to our expectation, we see higher correlations for stations separated along the crosswind direction.

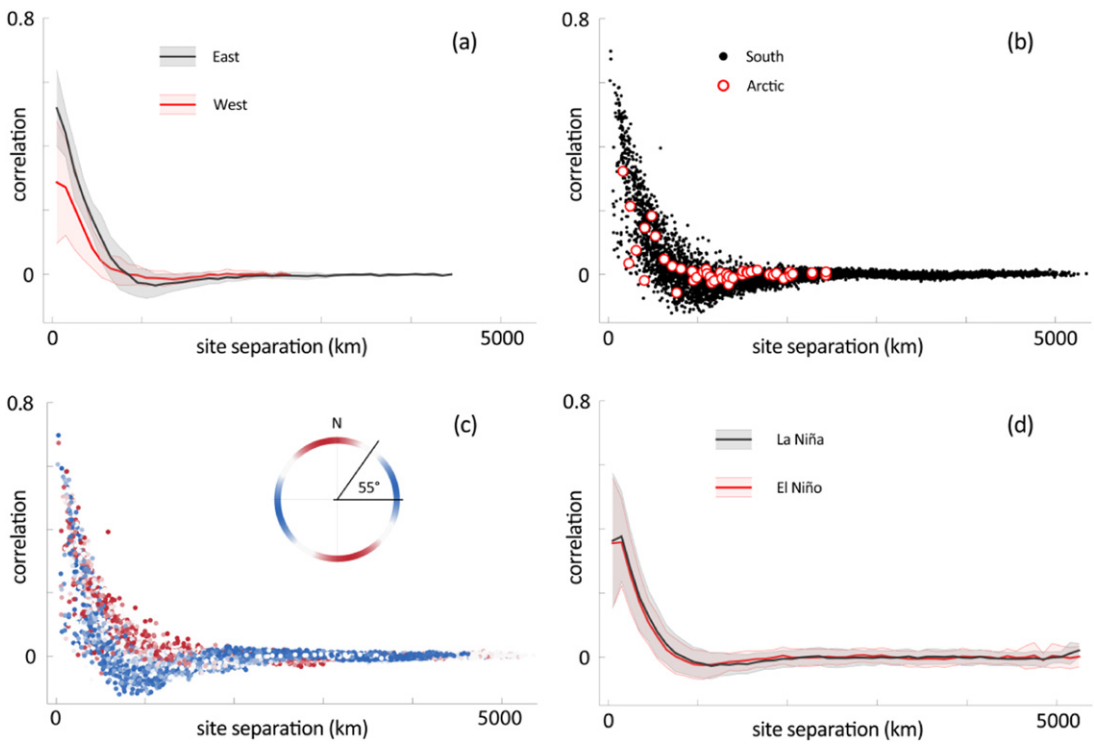

Figure 4. Correlation versus station separation for $\tau = 65$ h high-pass filtered CAN 1962–2006 hourly data. (a) Red: 1326 station pairs West of 100°W longitude, black 2080 station pairs East of 100°W longitude. Means of 100 km wide bins are shown by the solid line; the envelope represents ± one standard deviation σ; (b) red: 45 station pairs North of the Arctic Circle, black: 5671 station pairs South of the Arctic Circle; (c) azimuth of each station-pair bearing indicated by color in compass; (d) correlations during a 19 month period of strong El Niño (red) and during a 19 month period of strong La Niña (black). Means of 100 km wide bins are shown by the solid line; the envelope represents ± one standard deviation σ.

Finally, given previous work that has identified a connection between the interannual climate oscillation of the El Niño Southern Oscillation and Canadian wind resources (St. George and Wolfe 2009), we separated the time series into intervals with strong climate signals (1988–1989 La Niña and 1982–1983 El Niño, 1964–1965 La Niña and 1965–1966 El Niño, 1970–1971 La Niña and 1972–1973 El Niño), and calculated each site-pair's correlation coefficient over the El Niño interval and over the La Niña interval. Comparison of all these periods gave results similar to that seen in figure 4(d) for the 1964–1966 periods. No difference in correlations between El Niño and La Niña periods emerges (figure 4(d)). The pairwise differences between the '65–66 El Niño and the '64–65 La Niña correlations have mean $|\mu| = 0.0011$ and standard deviation 0.044 (see figure S8); the non-parametric sign test of null hypothesis H_0 that $\mu = 0$ against alternative hypothesis H_1 that $\mu \neq 0$, fails to reject H_0 with a p-value of 0.12.

3.2. Correlation length estimates

Figure 5 shows the results of our 'integral-scale' calculations for CAN, AUS and BPA. The 5300 km geographic extent of Canada clearly exceeds the observed correlation lengths, allowing the integration to fully saturate for even the widest filter, giving $\xi_{2160\mathrm{h}} = 273$ km. The AUS data yield slightly larger values, with $\xi_{2160\mathrm{h}} = 368$ km. The geographic extent of the BPA region is not sufficient for the correlation integration to saturate, making the $\xi_{2160\mathrm{h}} = 89$ km value, obtained an underestimate of the full correlation length.

Figure 6 shows the variation of correlation length ξ_τ with high-pass filter cutoff τ for the three regions, the ξ values for each region normalized by the maximum to facilitate comparison between datasets. The envelope surrounding the correlation length values as a function of τ indicates an empirically-found range of behavior common to all three datasets:

$$\frac{1}{\left[1 + \frac{38\ \mathrm{h}}{\tau}\right]^2} < \frac{\xi_\tau}{\xi_{\max}} < \frac{1}{\left[1 + \left(\frac{38\ \mathrm{h}}{\tau}\right)^2\right]^{1/2}}. \quad (2)$$

3.3. Discussion

The characteristics of the correlation length-scale metric introduced in equation (1) and used above deserve further explication. This measure has the advantage that it is essentially free of assumptions about the functional form of the correlation versus distance data. It quantifies the site-separation distance needed to reduce average inter-station correlation to a small value. However, it is important to remember, especially with data with such large scatter as observed here, that ξ_τ is

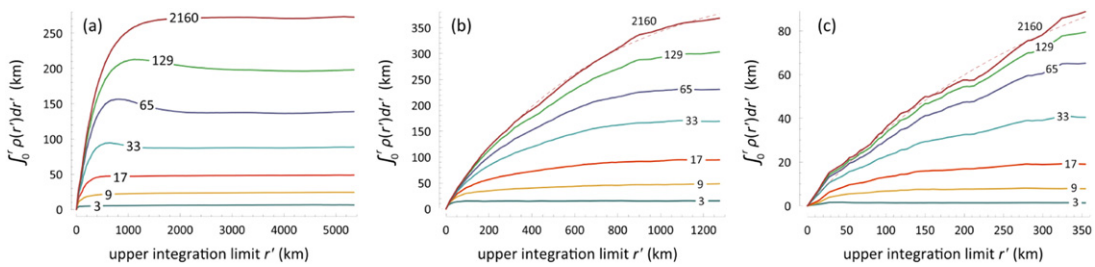

Figure 5. Numerically integrated correlation as a function of integration range for indicated high-pass filter window widths in hours for: (a) CAN, (b) AUS, and (c) BPA; 2160 h curves result from removal of diurnal cycle and 90 day moving average—curves with shorter window widths are further high-pass filtered. Dashed lines in (b) and (c) are least-squares fits to the form $\beta\ell[1-e^{-(r/\ell)}]$, giving, at $\tau = 2160$ h, $\beta = 0.65$ and $\ell = 685$ km for AUS, and $\beta = 0.40$ and $\ell = 323$ km for BPA. Similar fit gives $\beta = 0.70$ and $\ell = 388$ for CAN (not shown).

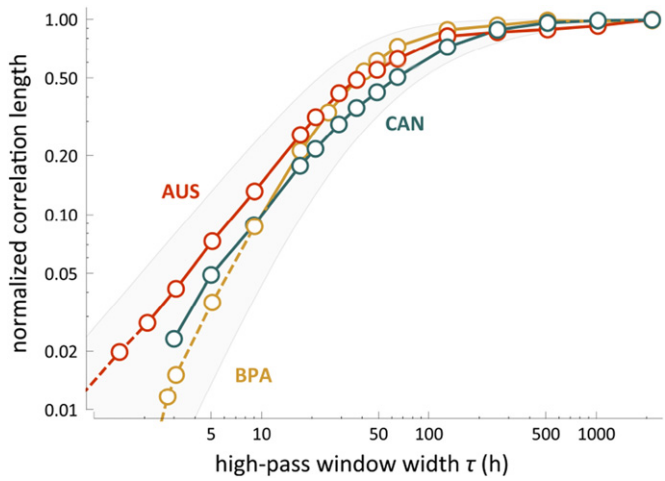

Figure 6. Correlation length versus high-pass window width for AUS, CAN and BPA data; shaded envelope spans $[1+\tau_0/\tau]^{-2}$ to $[1+(\tau_0/\tau)^2]^{-1/2}$, with $\tau_0 = 38$ h. Dashed curve segments dependent on non-zero nugget.

just one of multiple possible measures of 'how far is far enough'. For example, as seen in figure 1 for the CAN data, numerically integrating the scattered $\rho(r)$ points according to equation (1) gives $\xi_{65h} = 139$ km (right-hand end of light-blue curve), while integrating the dark-blue robust local-regression curve for the same window width gives the significantly larger value of 172 km (integral not shown). This occurs because the residuals of the local-regression fit are negatively skewed, and the 'robust' fitting process (Cleveland 1979) pulls the fit towards the median by assigning small weights to points determined to be outliers; here more often below the curve than above.

The integral-scale metric ξ_τ measures the separation distance required for correlation to fall to a value small compared to unity; this can be substantially less than the distance ℓ over which it falls to a fraction (say, $1/e$) of its initial value (at $r = 0$) if the initial value is small. This discrepancy between the integral-scale metric and the $1/e$ distance ℓ can be seen from the parameters of fitting of an exponential form to the data, as shown by the dashed-lines in figures 5(b) and (c). Here we fit $\xi(r) = \beta\ell(1-e^{-r/\ell})$ as would be the case for correlation falling as $\beta e^{-r/\ell}$ with distance.

Table 3. Comparison of least-squares fit parameters (β, ℓ) to integral-scale correlation-length metric (ξ_{max}).

Dataset	τ (hr)	β	ℓ (km)	$\beta\ell$ (km)	ξ_{max} (km)
CAN	2160	0.70	388	273	273
AUS	2160	0.65	685	447	368
BPA	2160	0.40	323	130	89

Fitting results are shown in table 3. Due to the large nugget effect (Matheron 1963) observed in all three regions, the best-fit correlation at zero site-separation β is substantially smaller than unity, proportionately reducing the area ξ under the $\rho(r)$ curve. For the CAN data this gives $\xi(r_N) \approx \beta\ell$, but in the AUS and BPA regions $\xi(r_N) < \beta\ell$ since the regions sizes do not permit separations r large enough to bring $\rho(r)$ to zero and hence to bring ξ to saturation.

The nugget effect has a particularly strong influence on ξ_τ values at small τ. Had we constrained $\rho(0) = 1$ rather than estimating the value at the origin using local regression, figure 6 would have shown the decline in ξ_τ/ξ_{2160} ending at a floor value limited by our trapezoid-rule integration to half the separation of the closest site pair (0.03, 0.01, and 0.07 for CAN, AUS, and BPA,

respectively), affecting only the dashed portion of the curves.

4. Conclusions

Variability in wind-generated electricity can be reduced by interconnecting wind farms across large regions, as distant wind farms are only weakly correlated. We investigate the geographic extent needed for this aggregation to be effective using three extensive data sets. Using a 'non-parametric' estimator ξ based on the area under the correlation versus distance curve, we found for the slowest variabilities $\xi_{max} = 273$ km for 10 m wind speeds in Canada, 368 km for wind-plant generation in Southeastern Australia, and 89 km for tower wind-speeds in the Northwestern US. Since the Australia and BPA regions are small enough that even for the most distant sites correlation never drops to zero, our ξ values for the widest filter window-widths are underestimates of the extent needed for fully effective aggregation. Quantities more representative of the extent needed for smoothing the slowest variabilities are given by the larger $\beta\ell$ values of 447 and 130 km respectively, in table 3, obtained by fitting $\xi(r)$ to the form $\beta\ell(1-e^{-r/\ell})$ and extrapolating ξ to $r=\infty$. These values can be compared to like values from previous work listed in table 1.

Although the regional length scales have different magnitudes, we find a dependence of ξ on variability time-scale that is remarkably similar across the three regions, as seen in figure 6. At time scales τ shorter than 38 h, ξ falls at least as fast as τ^{-1}, while at longer scales it is essentially constant. Thus, on time scales longer than a day or so, the variability-reduction benefit of aggregating wind power over a region of a given size will be independent of time scale. For time scales shorter than a day, the faster the variability, then the more smoothing that region could provide. It is the shrinking of correlation length with time scale that gives high-frequency spectral slopes a larger magnitude for power aggregated over a region compared to a single site (Beyer *et al* 1990, McNerney and Richardson 1992, Nanahara *et al* 2004, Katzenstein *et al* 2010, Tarroja *et al* 2011, Fertig *et al* 2012); similarly for solar power (Curtright and Apt 2008, Marcos *et al* 2011).

Our findings help disentangle the effects on variability reduction of *generator number, region size*, and variability *time scale*. In general, aggregating the outputs of N uncorrelated generators should reduce variability magnitudes by a factor of $\sqrt{N}$. A geographic region of area A has roughly $N \approx A/(2\xi)^2$ potential uncorrelated sites; thus, for a fully populated region (Hasche 2010) variability could be attenuated by a factor up to $\sqrt{A}/(2\xi)$. For time scales shorter than $\tau \approx 1.5$ days, the number of potential uncorrelated sites within a fixed area, and hence achievable attenuation, grows at least as fast as $1/\tau$. If the actual number of generators in a region is less than $A/(2\xi)^2$, the lesser number will determine attenuation.

Further work to better understand whether the high degree of scatter in correlation versus distance is a stable manifestation of some unidentified geographic process or is just persistent, random temporal variation that would average away over longer records would improve model utility. Additionally, analysis of solar data over a large region could determine if time and length scale of solar variability are linked in a way similar to what we found here for wind.

Acknowledgments

Wind data used in this work was provided by the Australian Energy Market Operator (AEMO), Environment Canada, and the Bonneville Power Administration, United States Department of Energy. This work was partially supported by a seed grant from Colorado Research and Education in Wind. This material is based upon work funded by the National Science Foundation under Grant IIP-1332147.

References

Achberger C, Chen D and Alexandersson H 2006 The surface winds of Sweden during 1999–2000 *Int. J. Climatol.* **26** 159–78

Adams T and Cadieux F 2009 Wind power in Ontario: quantifying the benefits of geographic diversity *Presented at the 2nd Climate Change Technology Conf. (Hamilton, Ontario)* pp 1–15

Australian Energy Market Operator 2013 *South Australian Wind Study Report* iv, B-2

Archer C L and Jacobson M Z 2003 Spatial and temporal distributions of US winds and wind power at 80 m derived from measurements *J. Geophys. Res.* **108** 4289

Archer C L and Jacobson M Z 2007 Supplying baseload power and reducing transmission requirements by interconnecting wind farms *J. Appl. Meteorol. Climatol.* **46** 1701–17

Baïle R and Muzy J F 2010 Spatial intermittency of surface layer wind fluctuations at mesoscale range *Phys. Rev. Lett.* **105** 1–4

Baïle R, Muzy J F and Poggi P 2011 Short-term forecasting of surface layer wind speed using a continuous random cascade model *Wind Energy* **14** 719–34

Batchelor G K 1953 *The Theory of Homogenous Turbulence* (Cambridge: Cambridge University Press) pp 47–8

Beyer H G, Luther J and Steinberger-Willms R 1990 Fluctuations in the combined power output from geographically distributed grid coupled wind energy conversion systems—an analysis in the frequency domain *Wind Eng.* **14** 179–92

Beyer H G, Luther J and Steinberger-Willms R 1993 Power fluctuations in spatially dispersed wind turbine systems *Sol. Energy* **50** 297–305

Bonneville Power Administration *Meteorological Information from BPA Weather Sites* (http://transmission.bpa.gov/business/operations/wind/MetData.aspx)

Bretherton C S, Widmann M, Dymnikov V P, Wallace J M and Bladé I 1999 The effective number of spatial degrees of freedom of a time-varying field *J. Clim.* **12** 1990–2009

Buell C E 1972 Correlation functions for wind and geopotential on isobaric surfaces *J. Appl. Meteorol.* **11** 51–9

Carlin J and Haslett J 1982 The probability distribution of wind power from a dispersed array of wind turbine generators *J. Appl. Meteorol.* **21** 303–13

Cleveland W S 1979 Robust locally weighted regression and smoothing scatterplots *J. Am. Stat. Assoc.* **74** 829–36

Cosseron A, Gunturu U B and Schlosser C A 2013 Characterization of the wind power resource in Europe and its intermittency *Energy Procedia* **40** 58–66

Curtright A E and Apt J 2008 The Character of power output from utility-scale photovoltaic systems *Prog. Photovolt.* **16** 241–7
Davenport A G 1961 The spectrum of horizontal gustiness near the ground in high winds *Q. J. R. Meteorol. Soc.* **87** 194–211
Ernst B, Wan Y H and Kirby B 1999 Short-term power fluctuation of wind turbines: analyzing data from the German 25 MW measurement program from the ancillary services viewpoint *National Renewable Energy Laboratory Report* NREL/CP-500-26722
Fertig E, Apt J, Jaramillo P and Katzenstein W 2012 The effect of long-distance interconnection on wind power variability *Environ. Res. Lett.* **7** 1–6
Fisher S M, Schoof J T, Lant C L and Therrell M D 2013 The effects of geographical distribution on the reliability of wind energy *Appl. Geography* **40** 83–9
Fitch A C, Lundquist J K and Olson J B 2013 Mesoscale influences of wind farms throughout a diurnal cycle *Mon. Weather Rev.* **141** 2173–98
Gibescu M, Ummels B C and Kling W L 2006 Statistical wind speed interpolation for simulating aggregated wind energy production under system studies *Presented at the 9th Int. Conf. on Probabilistic Methods Applied to Power Systems KTH (Stockholm, Sweden)* pp 1–7
Giebel G 2000 On the benefits of distributed generation of wind energy in Europe *PhD Thesis* Carl von Ossietzky Universität Oldenburg 1–104
Gunst R F 1995 Estimating spatial correlations from spatial-temporal meteorological data *J. Clim.* **8** 2454–70
Hasche B 2010 General statistics of geographically dispersed wind power *Wind Energy* **13** 773–84
Haslett J and Raftery A E 1989 Space-time modelling with long-memory dependence: assessing Ireland's wind power resource *Appl. Stat.* **38** 1–50
Hill D C, McMillan D, Bell K R W and Infield D 2012 Application of auto-regressive models to UK wind speed data for power system impact studies *IEEE Trans. Sustainable Energy* **3** 134–41
Hinkelman L M 2013 Differences between along-wind and cross-wind solar irradiance variability on small spatial scales *Sol. Energy* **88** 192–203
Hoff T E and Perez R 2010 Quantifying PV power output variability *Sol. Energy* **84** 1782–93
Holttinen H 2005 Hourly wind power variation in the Nordic countries *Wind Energy* **8** 173–95
Huang J, Lu X and McElroy M B 2014 Meteorologically defined limits to reduction in the variability of outputs from a coupled wind farm system in the central US *Renew. Energy* **62** 331–40
Julian P R and Thiebaux H J 1975 On some properties of correlation functions used in optimum interpolation schemes *Mon. Weather Rev.* **103** 605–16
Justus C G and Mikhail A S 1978 *Energy Statistics for Large Wind Turbine Arrays* (Atlanta, GA: Georgia Institute of Technology)
Kahn E 1979 The reliability of distributed wind generators *Electr. Power Syst. Res.* **2** 1–14
Katzenstein W, Fertig E and Apt J 2010 The variability of interconnected wind plants *Energy Policy* **38** 4400–10
Kempton W, Pimenta F M, Veron D E and Colle B A 2010 Electric power from offshore wind via synoptic-scale interconnection *Proc. Natl Acad. Sci. USA* **107** 7240–5
Kirby B 2004 *Frequency Regulation Basics and Trends* (Oak Ridge, TN: Oak Ridge National Laboratory) TM-2004/291
Kiss P and Jánosi I M 2008 Limitations of wind power availability over Europe: a conceptual study *Nonlinear Process. Geophys.* **15** 803–13
Landberg L 1997 Availability and variability of the European wind resource *Int. J. Sol. Energy* **18** 313–20
Landberg L 1999 Short-term prediction of the power production from wind farms *J. Wind Eng.* **80** 207–20
Lave M and Kleissl J 2013 Cloud speed impact on solar variability scaling—application to the wavelet variability model *Sol. Energy* **91** 11–21
Louie H 2014 Correlation and statistical characteristics of aggregate wind power in large transcontinental systems *Wind Energy* **17** 793–810
Marcos J, Marroyo L, Lorenzo E, Alvira D and Izco E 2011 From irradiance to output power fluctuations: the pv plant as a low pass filter *Prog. Photovolt., Res. Appl.* **19** 505–10
Marcos J, Marroyo L, Lorenzo E and García M 2012 Smoothing of PV power fluctuations by geographical dispersion *Prog. Photovolt., Res. Appl.* **20** 226–37
Matheron G 1963 Principles of geostatistics *Econ. Geol.* **58** 1246–66
McNerney G and Richardson R 1992 The statistical smoothing of power delivered to utilities by multiple wind turbines *IEEE Trans. Energy Convers.* **7** 644–7
Milligan M and Factor T 2000 Optimizing the geographic distribution of wind plants in iowa for maximum economic benefit and reliability *Wind Eng.* **24** 271–90
Mills A D and Wiser R H 2011 Implications of geographic diversity for short-term variability and predictability of solar power *2011 IEEE Power and Energy Society General Meeting (San Diego, CA, 24–29 July 2011)* pp 1–9
Molly J P 1977 Balancing power supply from wind energy converting systems *Wind Eng.* **1** 57–66
Nanahara T, Asari M, Sato T, Yamaguchi K, Shibata M and Maejima T 2004 Smoothing effects of distributed wind turbines: I. Coherence and smoothing effects at a wind farm *Wind Energy* **7** 61–74
Palutikof J P, Cook H F and Davies T D 1990 Effects of geographical dispersion on wind turbine performance in England: a simulation *Atmos. Environ.* **24A** 213–27
Ramanathan Y, Kulkarni P and Sikka D R 1973 On a study of winter season wind structure at 500 mb in the Indian region for use in objective analysis of the wind field *J. Appl. Meteorol.* **12** 977–83
Robeson S M and Shein K A 1997 Spatial coherence and decay of wind speed and power in the North-Central United States *Phys. Geography* **18** 479–95
Šaltytė Benth J and Šaltytė L 2011 Spatial-temporal model for wind speed in Lithuania *J. Appl. Stat.* **38** 1151–68
Şen Z 1989 Cumulative semivariogram models of regionalized variables *Math. Geol.* **21** 891–903
Şen Z and Şahin A D 1997 Regional assessment of wind power in Western Turkey by the cumulative semivariogram method *Renew. Energy* **12** 169–77
Simonsen T K and Stevens B G 2004 Regional wind energy analysis for the Central United States *Presented at Global Wind Power* pp 1–15
Sinden G 2007 Characteristics of the UK wind resource: long-term patterns and relationship to electricity demand *Energy Policy* **25** 112–27
Sørensen P, Cutululis N A, Vigueras-Rodríguez A, Madsen H, Pinson P, Jensen L E, Hjerrild J and Donovan M 2008 Modelling of power fluctuations from large offshore wind farms *Wind Energy* **11** 29–43
St George S and Wolfe S Z 2009 El Nino stills winter winds across the Southern Canadian Prairies *Geophys. Res. Lett.* **36** 1–5
Tarroja B, Mueller F, Eichman J D, Brouwer J and Samuelsen S 2011 Spatial and temporal analysis of electric wind generation intermittency and dynamics *Renew. Energy* **36** 3424–32
Thomas P H 1945 *Electric Power From the Wind* (New York: Federal Power Commission) pp 1–74
von Kármán T 1948 Progress in the statistical theory of turbulence *Proc. Natl Acad. Sci. USA* **34** 530–9
von Meier A 2006 *Electric Power Systems: A Conceptual Introduction* (Hoboken: Wiley)
Wan H, Wang X L and Swail V R 2010 Homogenization and trend analysis of Canadian near-surface wind speeds *J. Clim.* **23** 1209–25
Wiemken E, Beyer H G, Heydenreich W and Kiefer K 2001 Power characteristics of PV ensembles: experiences from the combined power production of 100 grid connected PV systems distributed over the area of Germany *Sol. Energy* **70** 513–8