Wind power simulation using Correlated Innovation Matrix and Wavelet Multi-Resolution Analysis approaches

Dougal McQueen, Alan Wood, and Allan Miller
Electric Power Engineering Centre, University of Canterbury, New Zealand.
Email: dougal.mcqueen@pg.canterbury.ac.nz

Abstract—To meet carbon emissions targets, increased demand, and replace retiring plant it will be necessary to construct new electricity generation plant in New Zealand. One of the least cost, and carbon neutral methods of generating electricity is wind power. The intermittent and variable nature of wind coupled with the passive reaction of wind turbines ensures that wind power requires scheduled and spinning reserves. Reserve requirements can be alleviated, to an extent, by distributing wind farms throughout the country, exploiting the diversity of wind. However, spatial diversification may increase costs through reducing economies of scale in wind farm construction and increasing transmission requirements. Transmission expansion has long lead times, and needs good models of an uncertain future mix of generation size, type, and location. There is insufficient measured wind power or wind speed data to assess the trade-offs for envisaged wind development scenarios hence a model of wind power is required. The model must be temporally and spatially congruent with respect to wind, demand, and other generation types. In this paper specific models are developed applying wind speed time-series derived from a Numerical Weather Prediction model, temporal imputation methods, transformation to power using wind farm power curves, and assumptions concerning electrical and operational efficiencies. The temporal imputation, or turbulence modeling, is achieved through two methods: a Correlated Innovation Matrix approach, and a Wavelet Multi-Resolution Analysis approach. The models are used to simulate power time-series for seven wind farms, and subsets of these used to assess a centralised scenario and a diversified scenario. Results are compared with aggregate measured power time-series demonstrating the benefit of spatial diversification and illustrating differences in the turbulence modeling approaches.

I. INTRODUCTION

Whether to meet requirements for carbons emissions reductions, increased demand, or to replace retiring plant there will be a need for additional electricity generation in New Zealand. The excellent wind resources in New Zealand ensure wind power is one of the least cost ways of generating electricity. The turbulent nature of the wind and the passive reaction of wind turbines means that wind power is intermittent and variable requiring scheduled and spinning reserves.

The spatial diversity of wind resources means that it is possible to alleviate some of the reserve requirements by distributing wind farms throughout the country. However, distributing wind farms may increase costs by reducing scales of economy in wind farm construction and increasing transmission requirements and losses. To assess the benefit of spatial diversification it is necessary to construct a model that can simulate the wind power time-series of each wind farm in a given scenario.

It is desired that the model simulates the dynamic power output for each wind farm in a manner that is spatially and temporally congruent. This is requisite as the simulated time-series will be combined with the time-series for other generation sources and demand so that aggregate loading on the transmission network and reserve requirements can be determined.

There are insufficient quality wind speed measurements to enable simulation of wind power time-series for all envisaged future wind farm development scenarios. Thus it is necessary to construct a model using wind speed time-series derived from a Numerical Weather Prediction model, scale the time-series to be equivalent to that which influences a wind farm, and then convert the wind speeds to power. The simulated wind power time-series are compared with measured values by aggregating two subsets to represent centralised and diversified scenarios.

II. DATA

This study uses the Model Output Statistics (MOS) from a Numerical Weather Prediction (NWP) model to characterise the coarse resolution variations in weather, wind speed measurements from meteorological masts to develop models of turbulence and to scale the MOS to be representative of wind farm sites, and wind power measurements for validation. The approximate locations of the NWP model grid points, meteorological mast locations, and wind farm locations are presented in Figure 1.

A. Synoptic wind speed

The coarse characterisation of wind speed time-series are obtained from reanalyses, or hindcasts, of the historic synoptic weather patterns performed by a Numerical Weather Prediction (NWP) model. The NWP model used is the European Centre for Medium Range Weather Forecasting (ECMWF) ERA-40 product [1]. The surface 10m wind speeds have been extracted for grid points covering New Zealand. The ERA-40 model has a spatial resolution of 0.5 x 0.5 degrees and a temporal resolution of 6h.

B. Wind speed

Wind speed measurements from twenty masts sited at prospective wind farm sites have been collated. This data set is broadly a duplicate of that used in the Wind Generation Integration Project [2]. A subset of the data is extracted spanning 248 days with missing values imputed using multiple linear regression. The wind speed time-series with the
most missing values requires 27% of the values imputed, however most of the time-series have no missing values. The wind speeds have been measured at a temporal resolution of 10 min.

The wind speed measurements are made at prospective wind farm sites, this ensures that the measurement locations are well exposed and that quality anemometry has been used, thus sites are not overly influenced by local topography and climatic conditions.

C. Wind power

Wind power time-series have been obtained from seven wind farms for a period of 184 days. The data represent total exported power and thus are subject to the vagaries of the power system and wind farm operation. Thus the wind power time-series include restrictions placed on exported power (run-backs), faults, operational outages, and electrical losses. To account for the restrictions data from SCADA has been obtained for each wind farm and operational efficiency time series calculated. Details of the wind farms are presented in Table I. The wind power time-series have a temporal resolution of 4 s.

### TABLE I

<table>
<thead>
<tr>
<th>Wind Farm</th>
<th>Capacity [MW]</th>
<th>No. of turbines</th>
<th>Proxy mast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Te Apiti</td>
<td>90.8</td>
<td>55</td>
<td>Tararua 1</td>
</tr>
<tr>
<td>Tararua 1.2</td>
<td>31.7</td>
<td>103</td>
<td>Tararua 2</td>
</tr>
<tr>
<td>Tararua 3</td>
<td>93</td>
<td>31</td>
<td>Tararua 3a</td>
</tr>
<tr>
<td>Te Rere Hau</td>
<td>48.5</td>
<td>97</td>
<td>Tararua 3b</td>
</tr>
<tr>
<td>West Wind</td>
<td>142.6</td>
<td>62</td>
<td>Wellington 1</td>
</tr>
<tr>
<td>White Hill</td>
<td>58</td>
<td>29</td>
<td>Southland 2</td>
</tr>
</tbody>
</table>

#### III. Wind speed time-series simulation

Wind farms in New Zealand typically consist of a group of propeller type turbines located on an exposed ridge or hill top. The wind that a wind farm experiences is dependent upon atmospheric forcing due to the passage of synoptic cyclones and anti-cyclones, local climatic conditions (such as atmospheric stability, and topography). A useful abstraction is to allow the processes involved in the progression of synoptic scale weather systems to be viewed as a deterministic process, usefully characterized by Numerical Weather Prediction (NWP) models that resolve the atmosphere by applying fluid dynamic parametrisations. And, the local climatic conditions to be related to turbulence and as such a stochastic process. Thus wind speed time-series can be generated by addition of the scaled output from a NWP model and the results from a turbulence model. Thus making the assumption that wind speed time-series are co-deterministic for frequencies greater than that equivalent to the temporal resolution of the NWP.

A. Spatial interpolation

The NWP model used here produces 10 m wind speeds on a regular 0.5 x 0.5 degree grid. To approximate the wind speed at each meteorological mast two dimensional spline interpolation is applied. The resulting wind speed time-series represents the wind flow at a height of 10 m free from the influence of local topographic features and turbulence.

B. Static scaling

To account for local topography the Measure Correlate Predict (MCP) algorithm is applied to each mast [3]. MCP requires solving a linear regression, forced through the origin, between coincident values of the NWP interpolated wind speed and the measured wind speeds for each of 12 wind direction sectors. Low wind speeds, less than 5 m s\(^{-1}\), are excluded in forming the linear regressions. The resulting ”speed ups” are then applied to the wind speed time-series derived from the NWP model.

C. Temporal imputation

The 6 hr resolution of the NWP model is insufficient for describing the variability of wind power; thus it is necessary to impute the time-series. Wind is turbulent and hence a stochastic process however it also has certain characteristics governed by the physics of air. The Kolmogorov spectrum describes the relative energy contained in vortices of different sizes, and is governed by the viscosity of air [4]. Thus it is possible to impute a wind speed time-series through the addition of coloured noise. However, the relationships typically used to generate coloured noise for sets of time-series that have significant correlation or interdependence, such as the Veers method [5], assume homoskedasticity and do not necessarily apply for scales of up to 1000 km. A numeric method that, to an extent, captures both the spectral power and interdependence is Vector Auto-Regression (VAR), an extension to the well known time-series methods described by Box-Jenkins [6]. However, VAR methods require a large number of parameters that are specific to the set of time-series used in derivation thus a generalisation is not easily constructed. A simplification of the VAR model is the Correlated Innovation Matrix (CIM) model which is applied here.
1) Correlated Innovation Matrix: The turbulence in the wind is only weakly related to the long term average wind speed \((R^2 < 0.15\) for all series), hence it is possible to identify a turbulence time-series using the relationship in Equation 1. Thus from the set of wind speed time-series a set of turbulence time-series is found and a CIM model formed to simulate these.

The CIM model uses a Box-Jenkins time-series approach; the incremental wind speed is related to a certain number of prior values plus a random innovation, where the number of prior values defines the model order. Using auto and partial-auto correlograms the model type and order can be determined. In the case of the turbulence time-series here, an Auto-Regressive model of order 3 is adopted. The correlation between turbulence time-series is fixed by weighting the innovations so that they are not independent. The CIM method is presented in Equation 2. The CIM method was adopted in the production of the New Zealand Synthetic Wind Data [7].

\[
\text{wind} = \pi + u_t'
\]  
\[
u_t = u_t + u_t'
\]

Where \(\text{wind}\) is the wind speed, \(\pi\) is the average wind speed, \(u_t'\) is the turbulent component of wind speed, and \(t\) is time.

\[
u_{t,n} = A_{t-1,n} \cdot \nu_{t-1,n} + A_{t-2,n} \cdot \nu_{t-2,n} + \cdots + A_{t-\alpha,n} \cdot \nu_{t-\alpha,n} + \zeta \cdot e_t
\]

Where \(A\) is the AR coefficient of order \(\alpha\), \(n\) is the mast index, \(\zeta\) is the CIM loci, and \(e\) is the innovation matrix.

The CIM method requires the simulated process to be stationary and Gaussian. To approximate stationarity the long term mean must not vary and the process must be homoskedastic. Here in order that the turbulence series can be considered stationary it is necessary to remove a diurnal cycle reflecting changes in atmospheric stability. Further shaping is not required as after the turbulence time-series are normalised they are found to be approximately Gaussian.

The weighted correlation matrix, or CIM loci, \(\zeta\) is found by forming a matrix where each element equals the Pearson’s correlation coefficient for pairs of time-series and taking the Cholesky decomposition. A generalisation of the correlation relationships is made by fitting a log linear function of the separation distance as shown in Figure 2.

Simulated turbulence time-series are produced by generating the appropriate length of Gaussian distributed random numbers. The set of numbers is multiplied by the weighted correlation matrix, and then each series iterated through its auto-regressive weighting and diurnal means are added. The turbulence time-series are then added to the scaled NWP model derived wind speed time-series to obtain a coherent set of 10 min resolution wind speed time-series.

The simulated wind speed time-series are compared with the wind-speed data set in Figure 3. The top left pane shows the model reasonably replicates the probability distributions. The top right pane presents the frequency spectra of the measured time-series (blue) and the simulated time-series (red). Frequencies less than \(4.6 \times 10^{-5}\) Hz are derived from the NWP, those greater than this frequency derive from the CIM model. Note the dip in the simulated spectra at approximately \(5 \times 10^{-5}\) Hz. The lower left pane presents the simulated Pearson’s correlation coefficient versus the measured for pairs of sites, the correlations are uniformly over predicted as a result of the NWP model’s physics. The lower right pane presents the mean error in the scaled correlation as a function of scale (blue), and root mean square scaled correlation error (red). Note the error in scaled correlations is low for periods less than a day.

2) Wavelet Multi-resolution Analysis: While the CIM method does capture most of the spectral characteristics observed in the measured wind speed time-series, it is seen that there is a dip in the frequency spectra at \(5 \times 10^{-5}\) Hz. It is desirable to seek a method that is able to capture the spectral characteristics, maintain the correlation structure, and allow for the heteroskedasticity of turbulence. There have been many methods applied to this problem including frequency dependent Fourier transforms [8], Copulas, and wavelet transforms [9], [10]. Here we develop an approach using Wavelet Multi-resolution Analysis (WMA).

Turbulence can be construed as a train of vortices with differing spatial sizes. The amount of time a vortex takes to pass over a point or plane is related to its physical size and the larger the vortex the more extensive its influence. Hence, it is reasonable to expect that the correlation between two
wind speed time-series will be more significant the lower
the frequency of fluctuations.

The quantity of information in a time-series is limited
by its resolution. A Fourier transformation converts a signal
into the frequency domain, using the sine function as its
basis, with frequencies spanning the fundamental through the
Nyquist frequencies. The total number of discrete frequencies
equals half the number of samples in the time-series, as
each component has both magnitude and phase. Hence there
is preservation of information. However, for each frequency
component, there is no information as to where in time
an oscillation occurs and each finite oscillation must be
“constructed” using a natural frequency plus complementary
frequencies necessary to cancel the oscillation at times when
it is not present. It is desirable to segment the time-frequency
plane in a manner that trades off the localization in time with
accuracy in the frequency of oscillations while preserving information in the signal.

Wavelets are compactly supported oscillatory functions
that are square integrable. A decomposition of a time-series
is performed using a fast wavelet transform by convolving the
wavelet across the time-series. The decomposition is
achieved recursively by dilating the wavelet. At each stage
of the decomposition a residual time-series and a wavelet-
series is obtained. A wavelet is defined in Equation 3 and
achieved recursively by dilating the wavelet. At each stage
of decomposition the resolution marries the

\[ \psi_{j,\tau}(t) = \frac{1}{\sqrt{2^j}} \psi\left( \frac{t - 2^j \tau}{2^j} \right) \]  

(3)

Where \( \psi \) is the mother wavelet, \( j \) is the scale, and \( \tau \) is the translation.

\[ \langle u(\tau, 2^j), \Psi(j, \tau) \rangle = \int_{-\infty}^{\infty} u(t) \frac{1}{\sqrt{2^j}} \psi^*(\frac{t - \tau}{j}) \, dt \]  

(4)

Where \( \Psi \) is the wavelet coefficient series, \( t \) is time, and * denotes the complex conjugate.

Because of the dyadic structure, the time-series are averaged
to a temporal resolution of 11.25 min such that after 5 stages of
decomposition the resolution marries the 6 h period of the NWP model. If the a wind speed time-series with a
resolution of 11.25 min is decomposed using 5 stages there
will result wavelet series with scales of 22.5, 45, 90, 180, and
360 minutes and a residual time-series with a resolution of 360 min. Care must also be taken as the non-symmetry of the
wavelet results in translation of the residual time-series; thus
the decomposition must be structured such that the residual
time-series are coincident with NWP model time-stamps.

The wavelet transform effectively increases the dimen-
sionality of the time-series, and this increases the complexity
of simulation. To reduce the dimensionality a wavelet shape
is selected such that the correlation between wavelet series
of adjacent scales is minimised. This selection may pose a
contradiction; either the wavelet shape mirrors some natural
oscillation and thus resolves the time-series into independent
time-frequency bins, or the wavelet shape is such that it
obfuscates the information in the time-series. For the wind
data set used in this study the Beylkin wavelet is found to
provide the lowest cross correlation. For further explanation
of the method please refer to [12].

The wavelet decomposition resolves each time-series into
5 wavelet coefficient \( \Psi \) series, thus at each scale there are
20 \( \Psi \) series (one for each meteorological mast) and these
are modeled as a set of time-series using a CIM approach.
so that correlations and auto-correlations are replicated. So
that the CIM method can be applied it is necessary to ensure
the \( \Psi \) series are homoskedastic and Gaussian.

It is found that the \( \Psi \) series are related to the mean wind
speed, or magnitude of the residual time-series, in contrast
to the assumption in Equation 1. A Taylor transformation
is applied as presented in Equation 5; the Taylor exponent
is found by optimisation such that the Pearson’s correlation
coefficient between \( u \) and \( |\Psi(j, \tau)| \), \( k = 1...K, j = J \)
is minimised.

\[ \Psi^{(T)}(j, \tau) = \frac{\Psi(j, \tau)}{u(\tau)^a} \]  

(5)

Where \( a \) is the Taylor exponent, and \( T \) represents the Taylor
transformed variable.

\[ \Psi^{(T)}(j, \tau) = \gamma + \eta \sinh^{-1}\left( \frac{\Psi^{(T)}(j, \tau) - \epsilon}{\lambda} \right) \]  

(6)

Where \( J \) represents the Johnson transformed variable, \( \gamma \)
and \( \eta \) are shape parameters, \( \epsilon \) is the location parameter, and \( \lambda \) is the scale parameter.

The Taylor transformed \( \Psi \) series are found not to be
Gaussian and hence a Johnson transform is applied. The
resulting \( \Psi \) series are found to be over-differenced and the
Box-Jenkins method suggest pure Auto-Regressive models
of order three. The CIM loci for each scale are found by
Cholesky decomposition of the pair-wise correlations. The correlations are found using log linear distance relationships
derived from the measured data, as presented in Figure 4.

Wind speed time-series are simulated by firstly replacing the
residual 6 h time-series, in the wavelet structure, with the
scaled NWP model output. Then for each scale of decompo-
sition, a set of normally distributed random numbers of the
appropriate length is generated. These are rank transformed and iterated through the auto-regression with the innovations multiplied by the appropriate CIM loci. The inverse of the Johnson and Taylor transforms are then applied to find the set wavelet coefficients. Once all wavelet series are simulated the inverse wavelet transform is processed to obtain the set of wind speed time-series with resolution 11.25 min.

The results are presented in Figure 5. The WMA method produces very similar results to the CIM method; note that the spectral dip in the amplitude spectra has disappeared, and the RMS error in scaled correlations for periods less than one day is lower than for the CIM method.

The low pass filter constant is necessary to account for the inertial and spatial averaging of a wind farm in comparison with that of an anemometer. It is found that a first order low pass filter is most appropriate, as proposed by proposed by Madsen [13], and described in Equation 7.

\[ u_\omega' = F'( \frac{F(u)}{1 - M\omega} ) \]  

(7)

Where \( \omega \) is frequency, \( F \) denotes the Fourier transform, \( F' \) the inverse Fourier transform, and \( M \) is the low pass filter constant.

B. Operational and electrical efficiency

The operational efficiency, often referred to as the availability, of a wind farm is affected by faults, maintenance outages, and curtailment of power exported to the grid or run-backs. Data for the operational efficiency is used to derive a Markov Chain model as described by Sulaiman [14]. Time series for the operational efficiency are simulated using this Markov Chain model and applied to the simulated wind farm power time series.

There is a lack of data and documentation for describing electrical efficiency which is assumed to equal 0.98.

V. SCENARIOS AND RESULTS

Using the wind speed to wind power transformation described in the previous section, wind power time-series are simulated for each wind farm coincident with the measured wind power time-series. From this set of wind farm power time-series two groups are constructed. The first, representing a centralised scenario, includes only wind farms in the Manuwatu, comprising Te Apiti, Tararua 1,2, & 3, and Te Rere Hau. The second, representing a diversified scenario, comprises White Hill, West Wind, and Te Apiti. Both scenarios aggregate to near 300 MW capacity, and the resulting power time-series have been normalised to this capacity to ensure parity.
The results of the wind power simulation using the CIM model are presented in Figure 7. The top left pane presents probability distribution functions. This shows the CIM model provides a good match to the measured values. And, the relative difference between centralised and diversified scenarios is captured by the simulations; the benefit of diversification is exhibited by a less curvaceous cumulative distribution function (CDF). The top right pane presents the probability of occurrence for ramp rates, positive and negative, over a 10 min time step. This shows the CIM model over predicts the ramp rates by approximately 50%. The lower left pane presents the 10th (dashed) and 90th (dotted) percentile ratios of predicted to simulated ramp rates as a function of lag. This shows the CIM model grossly over predicts ramps for short lags with the error decreasing markedly with increasing lag. The lower right pane presents the amplitude spectra, the spectra for the centralised scenario sits above the diversified scenario, note the dip in the spectra at $5 \times 10^{-5}$ Hz is evident however not as pronounced as for the wind speed simulations.

VI. CONCLUSION

Wind power time-series have been simulated using NWP model wind speed time-series interpolated and scaled to be representative of meteorological masts. The wind speed time-series have been imputed using two methods: CIM, and a WMA approach. The wind speed time-series have been compared with measured values and both methods shown to effectively replicate time-series statistics. The WMA method produces better frequency spectra characterisation than the CIM model which exhibits a dip in the amplitude spectra at approximately $5 \times 10^{-5}$ Hz.

The wind speed time series are converted to power by applying wind farm power curves, assumptions concerning operational and electrical efficiencies, and a first order low pass filter to account for inertial and spatial averaging. The wind farm power time-series have been grouped into two sets, one representing a centralised scenario and another diversified scenario. It is seen that the diversified scenario offers a flatter probability distribution, and lower ramp rates and that these properties are reproduced in the simulated time-series. It is seen that both models overestimate ramp rates as a result of the derived low pass filter constant being too small, however this overestimation allows differences in the CIM and WMA results to be observed.

ACKNOWLEDGMENT

This work is part of the GREEN Grid project funded by Ministry Business Innovation and Employment, New Zealand. www.epecentre.ac.nz/greengrid Thanks to the providers of data used in this study (Electricity Authority, Meridian, Trustpower, NZ Windfarms Ltd, Northpower, Wind Farm Group, Wind Flow, Wellington Regional Council, Unison).
REFERENCES


