Multi-year ten-minute synthetic wind speed time-series for 15 actual or proposed New Zealand wind farms.

NIWA Client Report: WLG2009-43
June 2009

NIWA Project: ELC09302

Multi-year ten-minute synthetic wind speed time-series for 15 actual or proposed New Zealand wind farms.

Richard Turner
Xiaogu Zheng
Neil Gordon (MetService)
Michael Uddstrom
Andrew Tait
Greg Pearson (MetService)
Rilke De Vos
Marina Sterk
Trevor Carey-Smith
Stuart Moore

NIWA contact/Corresponding author

Richard Turner

Prepared for

Electricity Commission

NIWA Client Report: WLG2009-43
June 2009
NIWA Project: ELC09302

National Institute of Water & Atmospheric Research Ltd
301 Evans Bay Parade, Greta Point, Wellington
Private Bag 14901, Kilbirnie, Wellington, New Zealand
Phone +64-4-386 0300, Fax +64-4-386 0574
www.niwa.co.nz

Contents

Reviewed by: *Approved for release by*

Dr Mike Revell **Dr Andrew Laing**

Executive Summary

The Electricity Commission has engaged NIWA (with MetService as a sub-contractor) to create a multi-year (five) synthetic wind dataset (SWD) for 15 actual or potential wind farm sites from different parts of the country for use in modelling the electricity system. The required properties of the SWD are demanding and not able to be met by available observed datasets. The Electricity Commission will make the SWD and this report publicly available for use by any interested party, as well as for its own purposes.

As a result of the efforts described in this report a five-year MM5-based 10-minute SWD and a two-year NZLAM-based 10-minute SWD have been created that are suitable for use in modelling wind farm impacts on the New Zealand electricity system. Histograms of wind-speeds and high frequency fluctuations at the 15 different sites have been reproduced very well for eleven of the fifteen sites, and reasonably well for the other four sites. Particular attention was paid to re-producing the tails of the histograms so that the frequencies of wind speeds outside the turbine operating ranges were accurately simulated. Attention was also paid to reproducing the observed inter-site correlations that result from the passage of frontal systems which are often associated with important rapid region-wide increases and/or decreases in wind speeds. Each site’s diurnal cycle was calculated in such a way as to reproduce very closely the observed diurnal cycle, but with realistic daily and seasonal modulation of the amplitude maintained. Inter-site correlations of high frequency (hour and ten-minute) fluctuations of wind speed were reproduced well.

The two SWD produced have no missing data and are in a format suitable for use in wind-generation calculation programs.

1. Introduction

1.1 Background

Wind data at ten-minute to hourly time scales is an important input into modelling the performance of wind farms and their impact on the national electricity system. However, relatively little ten-minute wind data at actual wind farm or proposed wind farm sites is publicly available in New Zealand. NIWA and MetService do collect ten-minute wind data at about 60 sites which are publicly available on NIWA's National Climate Database (CLIDB). However, these tend to be low-elevation sites and are generally available only within the past few years. Also the NIWA and MetService data are typically from 10 m masts, which is well below the hub height for turbines and lower than the mast data collected at wind farm sites.

In the absence of actual observed data, simulated or "synthetic" data can be used, providing it is well calibrated against real data, to produce realistic results for applications such as (i) creating generation scenarios during the passage of frontal systems, (ii) calculating the capacity contribution of wind, or (iii) estimating the contribution of wind during different seasons.

The Electricity Commission therefore has contracted NIWA (with MetService as a sub-contractor) to create a multi-year (five) synthetic wind dataset (SWD) for 15 actual or potential wind farm sites from different parts of the country for use in modelling the electricity system. The required properties of the SWD were demanding and are outlined below (Section 1.2). The Electricity Commission intends that the SWD and this report will be made publicly available for use by any interested party, as well as for its own purposes.

This report (i) describes the data and the method used to create the SWD (Section 2), (ii) compares the SWD against "observed" mast data from wind farms (Section 3), and (iii) discusses the results and draws conclusions (Section 4)

1.2 Required physical and statistical properties for synthetic wind time series.

The Commission requirements for the synthetic wind time series were that it:

- be valid for turbine altitudes
- include multiple years
- cover 15 sites corresponding to either existing, proposed or potential wind farms, with records time-stamped to preserve realistic meteorological inter-area correlations and allow modelling of the impact of specific meteorological events.
- reproduce well seasonal and synoptic variations in wind speed and preserve inter-site correlations
- reproduce well the higher frequency (hour to hour) and (ten-minute to ten-minute) fluctuations in wind speed and preserve inter-site correlations
- reproduce reasonably (given the constraints of “disguised” data and the fact power generated is proportional to the cube of the wind speed) the wind speed climatology for each site, i.e., the approximate wind resource, the histogram of wind speeds including the frequency of speeds below and above the turbine operating thresholds
- be at intervals of 10 minutes with no missing gaps
- be in a form that can be converted into simulated wind farm output

The Commission was not interested in accurately reproducing the relative wind-resources between sites, i.e., this study is not about identifying rich wind resource areas.

As a result of the efforts of NIWA and MetService within this project a continuous 10 minute SWD was produced for 15 sites at a hub height of 85m for both (i) a 5 year period based on an archive of hourly MM5 (MetService^{TM}) forecast wind, and (ii) a 2 year period based on an archive of hourly NZLAM (NIWA) forecast winds. Figure 1 shows the regions within which these 15 sites were located. As will be shown later in this report we demonstrate both the 5 year SWD and 2 year SWD meet fully all these requirements.

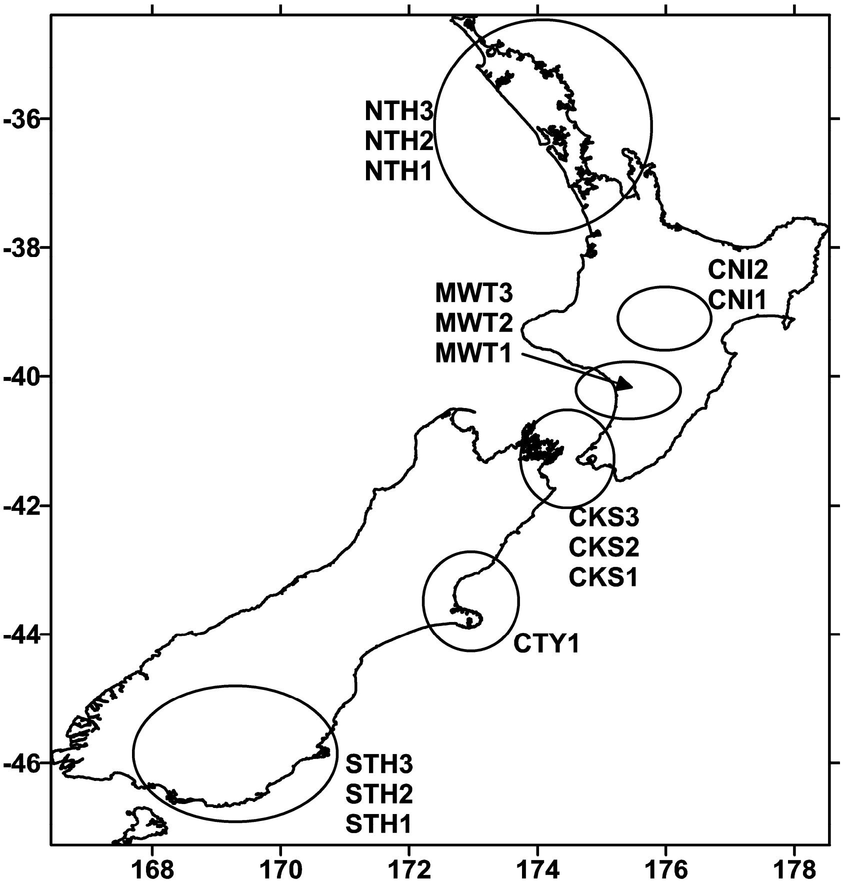


Figure 1 **The regions within which the wind farm sites were located for the SWD project. Note, STH is the regional identifier for Southland and Otago, CTY for Canterbury, CKS for Cook Strait, MWT for Manawatu and Wanganui, CNI for Central North Island and Hawkes Bay, and NTH Coastal parts of Waikato, Auckland, Coromandel, and Northland.**

2. Methodology

The philosophy behind the joint NIWA-MetService^{TM} approach was to utilise several years of archived 12 km gridded hourly NWP winds over the whole of New Zealand. These NWP winds have inter-site correlations and temporal variability (on a synoptic time-scale) that match the observations well. If a robust relationship between these hourly NWP winds and hourly speeds observed at wind farms could be developed, then imposing a high-frequency (ten-minute) residual that was uncorrelated between sites (as is observed, other than for sites that are less than a few km apart) but with the correct distribution of timestep to timestep fluctuations would produce a SWD with the desired properties. As will be seen there was some variation from this initial approach as the most robust relationships between modelled and observed relationships was obtained at daily time-scales. In this section, the data used and the NWP outputs are first described; this is then followed by details of the statistical methodology used.

2.1 Observed mast data – wind farm sites

The observed mast data provided by generators is commercially sensitive so for most sites it was disguised before being provided by either (i) normalising or (ii) not revealing the mast height and/or co-ordinates. The provided wind speeds were multiplied by some factor to ensure the average of observed data was a reasonable estimate for that site; these estimates were made to the nearest 0.5 m/s. The definition of reasonable estimate was based on our own knowledge of likely local terrain impacts on regional wind as represented by average speeds from nearby lowland climate stations, an annual average wind speed map produced by NIWA, and/or the NWP regional average. A summary of what was provided is given in Table 1.

Table 1: Facts about observed data provided for each site.

Site	Obs. Period	Disguise	% Missing	Disguised Mean Speed (m/s) at 85 m
STH1	Sep 2006-Aug 2007	Mast height not provided	0.7	10.0
STH2	Jan 2007-Dec 2007	Normalised	2.4	9.5
STH3	Jan 2007-Dec 2007	Mast height not provided	26.0	11.0
STH3A	Jan 2006-Dec 2006	Normalised	2.7	11.0
CTY1	Aug 2007-Jul 2008	Normalised	0.0	8.0
CKS1	Jan 2004-Dec 2004	Normalised/Composite	0.0/100	9.5
CKS2	Nov 2007 - Oct 2008	Mast height not provided	3.9	10.0
CKS3	Jun 2007-May 2008	Extrapolated from 10 m	4.1	11.0
MWT1	Jun 2007-May 2008	Normalised	0.8	10.0
MWT2	Aug 2006 – Jul 2007	Mast height not provided	2.5	10.0
MWT3	Aug 2007 – Jul 2008	Mast height not provided	0.0	9.0
CNI1	Aug 2007 – Jul 2008	Normalised	10.2	9.0
CNI2	Jan 2007 – Dec 2007	Mast height not provided	1.2	9.5
NTH1	Nov 2006 -Oct 2007	Mast height not provided	0.0	8.5
NTH2	Aug 2007 – Jul 2008	Mast height not provided	6.8	8.5
NTH3	Sep 2003-Dec 2008	Zero obs, composite of nearby climate network stations	100.0	8.5

2.2 Observed 10 metre data – CLIDB

In order to fill the small number of missing hourly wind speeds from the NZLAM archive, hourly surface (pulse anemometers on 10 m masts) winds from nearby lowland climate stations were extracted from CLIDB and extrapolated to 85 m using a power law. The alpha parameters used in this extrapolation, and the extrapolation of the NZLAM 10 m winds to 85 m were based on representative land use/vegetation type surrounding the sites. For site CKS1 there was no period where available observations overlapped with NZLAM; here climate station data from Baring Head and Cape Campbell were used to provide a composite dummy station with hourly records. Similarly data from Mokohinau and Musick Point were composited to construct a dummy observed time-series for site NTH3, as no observations from that possible wind farm site exist. Climate station data from Palmerston North and Dannevirke were also used in assessing the possible use of a method described by Haslett and Raftery (1998) (See Appendix A).

2.3 MM5 (MetService^{TM})

The wind archive data has been derived by running the local area model MM5 from 2003. MM5 was developed at Pennsylvania State University and the National Center for Atmospheric Research as a community based model.

The configuration used for the wind archive is a nested configuration with outer nest set to 36 km horizontal resolution (110x110 grid points) and inner nest set to 12 km (109x141 grid points). Every third grid point of the inner nest is coincident with a grid point of the outer nest. Note that the inner nest is not centred within the outer nest (see Figure 2). The horizontal grid has an Arakawa-Lamb B-staggering of the velocity variables, with scalar quantities defined in the centre of each grid square. The rest of the configuration is given in Table 2.

The vertical structure is a terrain following sigma coordinate, where sigma is defined

$$\sigma = (p - p_{\mathrm{top}})/(p_{\mathrm{surface}} - p_{\mathrm{top}}).$$

So σ is 1 at the surface and 0 at the model top which is set to 100 hPa, and there are 25 vertical levels.

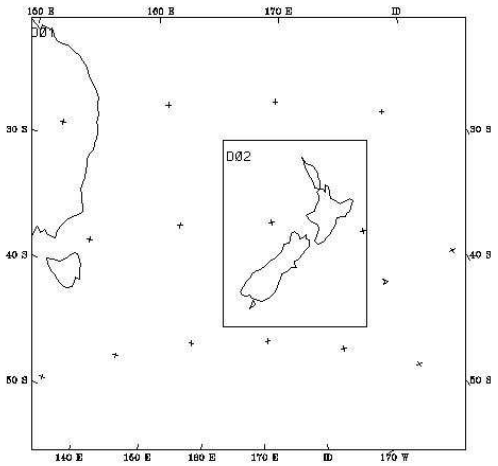


Figure 2. **Nesting structure of the MM5 configuration.**

The MM5 configuration was initialized and had the lateral boundary of the outer nest set by the NCEP GFS global model at 1 degree horizontal resolution. Where there were gaps in the archived data, these were filled by initialising the same MM5 configuration with the NCEP reanalysis data set at 2.5 degree horizontal resolution. MM5 was simply run in a ‘cold start’ fashion without any data assimilation. The model was run four times per day at initialization times of 0000 UTC, 0600 UTC, 1200 UTC and 1800 UTC. Other important MM5 settings are given in Table 2.

The model data extracted was the 3-dimensional grid of hourly winds over the inner nest, at prognosis times of T+6, T+7, T+8, T+9, T+10, T+11. These prognosis times were chosen so as to be far enough into the model run to avoid problems with model ‘spin up’. The extra prognosis time of T+12 was added in case it was needed to help alleviate discontinuity problems between successive model runs. Overall the MM5 winds were made available for this project from September 2003 through to August 2008. The NCAR software RIP (Read Interpolate Plot) was used to read the model data and interpolate it to various heights above the surface. The resulting heights were 10 m, 15 m, 30 m, 45 m, 60 m, 75 m, 100 m, 120 m, 150 m and 200 m. No horizontal interpolation was carried out, and the four surrounding grid points for each of the sites were provided.

Spectral filtering of the time-series was necessary to remove unphysical spikes in the extracted hourly time-series. These spikes were due to the discontinuity introduced by using the winds from T+6 to T+11 hours from consecutive forecasts; i.e., the winds at T+12 of one forecast are not always equal to the T+6 winds from the next. Also noise perhaps introduced through the assimilation scheme within the GFS forecast cycles would contribute to spikes at sub-6 hourly harmonics. Figure 3 shows the impact of the spectral filtering on the MM5 power spectra and Figure 4, while only a short segment of the overall time-series, shows the small impact this filtering had on the time series of NWP hourly wind speeds.

Table 2: A list of important parameter settings for the MM5 forecasts used in this study.

Model Characteristic	Type
Governing Equations	Non-hydrostatic (Grell et al, 1995)
Map Projection	Lambert Conformal
Nesting	Two-way (interactive)
Time Step	36 sec (outer nest 108 sec)
Initialisation and lateral boundary	1 degree resolution GFS
Cumulus Scheme	Grell (1994) – both nests
Planetary Boundary Layer Scheme	MRF (Hong, Pan 1996)
Microphysics Scheme	Mixed Phase (Reisner et al, 1998)
Radiation Scheme	'Cloud' – called every 30min
Orography	USGS 30 second resolution DEM
Vertical Resolution	25 levels concentrated near the surface
Model Top	100 hPa
Horizontal Resolution	12 km inner nest (36 km outer nest).
Vertical sigma levels	1.00, 0.998, 0.995, 0.99, 0.98, 0.96, 0.93, 0.89, 0.85, 0.80, 0.75, 0.70, 0.65, 0.60, 0.55, 0.50, 0.45, 0.40, 0.35, 0.30, 0.25, 0.20, 0.15, 0.10, 0.05, 0.00

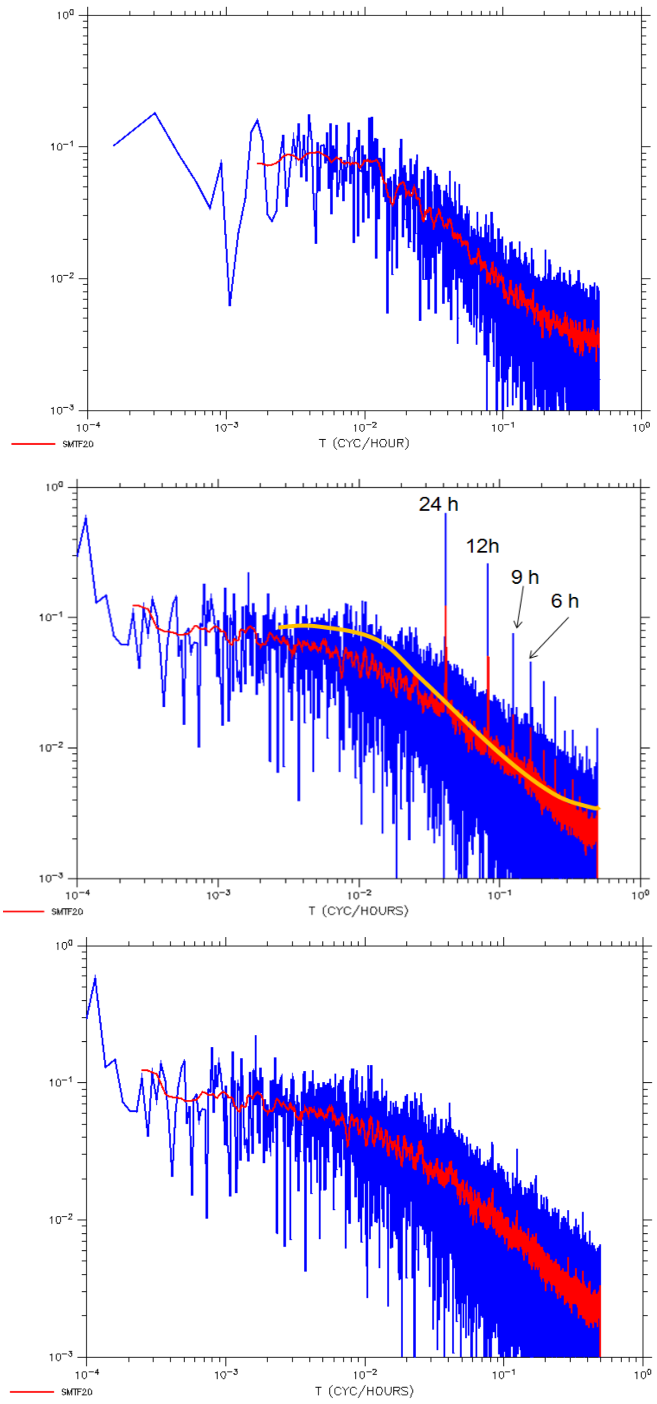


Figure 3: **Power spectra for hourly time series of wind speeds at site CKS3 as calculated from (a) 1 year observed (10 m mast) (b) 5 years of unfiltered MM5 (10 m winds NW grid point closest) hourly series, and (c) 5 years of filtered MM5 hourly series. Note the orange curve in (b) is the average slope of the observed spectra.**

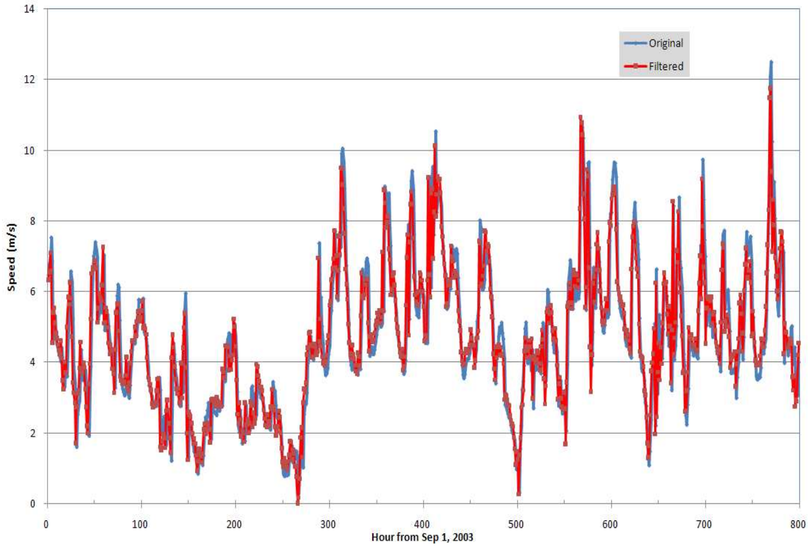


Figure 4 **Actual re-calculated time series of unfiltered and filtered hourly time series of MM5 10 m wind speeds from closest NW grid point at site CKS3 for the first 800 hours from 0600 UTC 1 September 2003**

2.4 NZLAM (NIWA)

NIWA's advanced weather prediction model, the New Zealand Limited Area Model (NZLAM) runs on a rotated latitude-longitude map projection, with the equator centred on the geographical domain of interest, leading to a nearly conformal 12 km grid mapping for its 324 × 324 × 38 (level) grid. The Met Office UK Unified Model or UM (of which NZLAM is a local implementation) is a non-hydrostatic, fully compressible, deep atmosphere formulation using a terrain-following height based coordinate. It uses a semi-Lagrangian advection scheme for prognostic variables and an Eulerian treatment of the continuity equation for mass conservation. It is run on a 6 hour analysis / forecast cycle, and two forecasts (at 0600 and 1800 UTC) each day are run to 48 hours (the other two run to the next analysis time window, i.e. 6 hours).

NZLAM is unique in that it includes the capability to assimilate local data. The data assimilation system is 3DVAR+FGAT (Lorenc et al, 2000), and the increments are estimated at the time of observation using a 3 hour data cut off, then added to the background using an incremental analysis update technique (Bloom et al, 1996), beginning 3 hours before the nominal analysis time and extending over 6 hours (i.e. to 3 hours after nominal analysis time). Data assimilated include standard meteorological observations from land and ocean stations (ships and buoys), rawindsondes and aircraft, as well as satellite observations of atmospheric radiances (HIRS, AMSU-A, AMSU-B), ocean surface wind speed and total column moisture (SSM/I), ocean surface wind speed and direction (ASCAT, QuikSCAT) and atmospheric motion vectors from geostationary meteorological satellites (MTSAT). Because this forecast system assimilates local data it can be "warm cycled". Accordingly, it does not utilise global model data within the forecast domain and accordingly preserves mesoscale structure between forecast cycles, with the result that NZLAM forecasts maintain their accuracy throughout the forecast period (i.e., they do not suffer from "spin up" problems like those that occur in non data assimilating models).

Where in situ data are available, these can be used to inform an adaptive statistical post processing system that removes site specific biases from forecast fields. Based on current results, wind speed forecast accuracy (at 10 m) is of the order of 1.5 to 2 ms^{-1} (evaluated at hourly intervals and all valid forecasts, i.e., hourly to 48 hours).

A comprehensive description of the Unified Model can be found in Staniforth et al (2004) and applications of the UM to simulating strong wind events in New Zealand are reported in Webster et al (2008). Some of the important model configurations and parameter settings of NZLAM are given in Table 3.

For this study surface (other model levels from within the PBL were not archived during this time) wind speed and direction output were extracted from the archives for forecast hours +0 to +12 for each forecast made between 24 October 2006 and 15 December 2008. The surface speeds were interpolated to approximate (the co-ordinates of the mast locations were typically not divulged) locations of the wind farm site then extrapolated to 85 m using a power law. In this 2 year period there were 4 short (i.e., 12 or 24) hour periods in April and May 2007 where no NZLAM output was available. For these periods, nearby surface hourly observations extrapolated from 10 m to 85 m were used to fill the gaps. Some spectral filtering of the time-series was also necessary to remove unphysical spikes in the extracted hourly time-series. In hindsight, given the fact that +0 to +3 forecasts are within the assimilation window, using +3 to +15 hour forecasts may have resulted in a smoother power spectra in less need of filtering.

Table 3: **A list of important parameter settings for the NZLAM forecasts used in this study.**

Parameter	NZLAM 12 km Setting
Domain size nx,ny,nz	324x324x38
Time step	300 s
Radiation time step	3600 s
No. levels in boundary layer (i.e., altitudes less than approx 1500 m)	13
Radiation scheme	3A (General 2-stream, see Edwards et al, 2004)
No. of segments	2
No. of bands	5
Convection scheme	4A (penetrative mass flux scheme,
	see Maidens, et al, 2004)
Includes momentum transport	Yes
Diffusion scheme	2A (explicit, see Staniforth et al, 2004)
Horizontal diffusion	Standard
Vertical Diffusion	Off
Diffusion coeffs for wind, theta, and moisture	K=-1.0, N=0

2.5 Detailed steps of the Statistical-NWP based method

The steps to our eventual method, developed after some trial-and-error, are described in this sub-section. For completeness, the first 3 steps which involve preparing the observed and NWP hourly data, while mentioned previously, are included in the outline below.

For each site:

Step 1: Extract the historical NWP (either MM5 or NZLAM) model output at hourly intervals for the available levels and for the four grid-points surrounding the site. Either horizontally interpolate (NZLAM) to the wind farm site or select one of the surrounding grid points (MM5) and then extrapolate (NZLAM) or interpolate (MM5) to the hub height of 85 m.

Step 2: Filter the 85 m NWP hourly time series so as to remove 6 hourly discontinuity and other unphysical spikes in the power spectra.

Step 3: Collect observed mast data (which is disguised – either it was normalised or the mast height and co-ordinates were not revealed) then (i) insert missing data flags, (ii) multiply data to ensure average of observed data was a reasonable estimate for that site, these estimates were to the nearest 0.5 m/s, and (iii) convert times to UTC.

Step 4: Derive a relationship which will use these filtered hourly NWP values as predictors to obtain ten-minute winds. The relationship is assumed to be of the form

$$w = a + bx + \sum_{i=-3}^{1} c_i y_i + d_1 \,|\cos(z)|\, + d_2 \,|\cos(z + 45^o)|\, + \varepsilon \qquad (1)$$

where *w* is observed 10 minute wind speed, *x* is the 24-hour running mean (centred) of modelled NWP (either MM5 or NZLAM) hourly wind speed *y* is the departure of modelled hourly wind speed from *x*, y_i is the *i*-hour lagged *y* (y_0=*y*), *z* is the wind direction in degrees (*z*=0 is for northerly). All NWP wind components are at hourly time steps, but these are linearly interpolated to 10 minute time steps. Finally, *a*, *b*, $c_{i\,(i=-3,-2,-1,0,1)}$, d_1 and d_2 are the parameters to be determined as the NWP winds are fitted to the data such as to maximize the explained variance and/or minimize the Bayesian Information Criteria (BIC) over the observing period. Values for all these parameters are given in Appendix B. Note, these parameters are unique for each site and the time period of the study. Finally, ε is the error.

Fitting the model to data (Step 4) was an involved process and the following notes elaborate on this process.

- x was chosen as 0, 24 and 48 hourly running mean of MM5 wind speed and was used to predict *w*. Using the 24 hourly running mean resulted in the lowest BIC, and therefore is preferred.

- The lagged and lead time series of *y* (modelled variability at hourly time scale) have been used as explanatory variables. Lags -3, -2, -1, and +1 are significant using BIC. The existence of significant lagged predictors is due to the fact that forecast (i.e., NWP) changes in wind speeds are often either earlier or later than predicted.

- For some sites, notably CKS3, MM5 seemed to under-estimate both southerlies and northerlies and especially southerlies. This was true for both “land” and “sea”. So $|\cos(z+45)|$ was chosen as a predictor and found to be statistically significant. This was confirmed to be significant for CKS1 as well. Values other than 45 were tested, but resulted in higher BIC values. Doing this improved the prediction variance (R-squared) for these sites; e.g., for CKS3, the R-Squared was 0.3366 for MM5 hourly prediction but 0.4353 following these steps.

- Apply the daily moving average filter to separate the error into the daily component and 10 minute component. Both component time series are assumed to be ARMA (Autoregressive-Moving Average) processes. The BIC is used to determine the orders of the autoregression and the moving average. For the both time series, the order for the autoregression is around 6 (corresponding to 6 ten minute periods in one hour) and the order for the moving average is around 2. The function “arima” in the statistical package R is used to fit Eq. 1 to data. This is critical for obtaining realistic distributions (histograms) of the time-series deltas (i.e., speed_{t+1} - speed_t).

- To ensure the variance of the simulated daily residual time series is close to that observed, the innovation error variance of the daily component time series is factored by the ratio of variance of observed daily residuals over the variance of the simulated daily residuals using the initially estimated parameters. For some stations, the innovation standard deviation of the 10 minute residual was a linear function of the NWP wind speed. This linear function is estimated by MLE (Maximum Likelihood Estimation).

- The function “arima.sim; is used for simulating the daily and ten minute residual time series. The innovation error for the 10 minute residuals is rescaled by the linear function of model prediction. The simulated daily error and 10 minute error are added onto the trend (estimated using Eq. 1) to form the simulated wind speeds. Negative wind speeds can be simulated due to the Gaussian error assumption in the AR models. These were re-distributed within the light wind (< 5 m/s for most sites, < 9.5 m/s for 3 of the windier sites) to more closely match the observed distribution. This re-distribution can be expressed mathematically as

  $$F_s(w)=F_o(w') \qquad (2)$$

  where F_s and F_o are the PDF’s of simulated (observed) wind speed below the 5 m/s threshold, and w is the simulated wind speed and can be redistributed to w' such that Eq. 2 is satisfied since Fs, Fo, and w are known.

  Plots, not shown, of the distribution of simulated 10 minute wind speeds generated using an assumption that the standard deviation of the innovation error was linearly related to the NWP wind speed showed a clear improvement over the simulated series where a constant innovation error was assumed.

- As will be seen in the results (Section 3.6) the observed diurnal range was generally larger than that simulated. To investigate the impact of the observed diurnal range, we removed it from the observation and repeated the previous analysis. However, this resulted in a very long range autocorrelation for the error ε. The long range correlation could be due to the existence of diurnal range in the NWP prediction, indicating that diurnal range should not be ignored. To further investigate this we split each 24 hour period into night (0000-0600 NZST), morning (0600-1200 NZST), afternoon (1200-1800 NZST) and evening (1800-2400 NZST) periods and fitted the wind speed for each of these periods. It shows beside the intercept (which is well related to diurnal range), that the other parameters were not sensitive to the period. To further test the impact of diurnal range, we modified our statistical model Eq. 1 to allow the intercept *a* to depend on the time of day. However, a larger BIC is produced for this modified Eq (1). Given that only one site (CKS1) had a diurnal amplitude > 1.5 m/s, we decided against using this modified model.

- Using model hourly data from all 4 surrounding NWP grid points as predictors was tested for site CKS3 against just using an interpolated (weighted by inverse distance) average or the grid point that correlated best with observations. A lower BIC resulted from using the latter.

- We experimented with “log” and “sqrt” transforms of *w*. However, the transformations over-simulated (i.e., the incidence of high wind speeds was far too high) the tail of the wind-speed distribution, especially for the log transformation. Experiments with a General Linear Model (GLM) with Poisson family led to a similar conclusion. We also tried to fit a linear predictor using a “gam” (Generalised Additive Model). However, the data seems to better fit a linear model which matches our physical understanding of the data and NWP models.

Once all these steps had been taken we had provisional ten-minute SWD for all 15 sites.

Inter-site relationships;

Inter-site comparisons were then done to check whether spatial correlations of the original NWP hourly wind speeds had been maintained. In general these inter-site correlations (see Section 3.8) were a bit lower than desirable, so an additional adjustment (Step 5) was made.

Step 5: Correlations between site pairs of wind speeds were generally a bit lower than observed. To correct for this we assumed the daily component of the residuals are correlated but the ten-minute component residuals are independent. The spatial dependence was then modelled by introducing a correlated Gaussian innovation error. The marginal Gaussian error was then redistributed to the fitted innovation distribution using a similar process as in Eq. 2. In other words, we choose a spatial correlation between the two daily residual time series such that the correlation between the two series using this correlation was close to that observed. Then we rescaled the ten-minute variance to get a better match of the tails, using the fact that the ten-minute variance is correlated with the wind speeds, i.e, for larger predicted wind speeds, the ten-minute variance was larger.

Step 6: In order to remove some artificial coarseness in the low-wind speeds introduced when the adjustment to correct the low speed tail of the site-histograms (Eq. 2) was made, the wind speeds within each bin of F_o were simply randomly re-distributed within that bin.

Step 7: In order to get a more realistic match of SWD diurnal cycle we calculated the difference between the observed average diurnal cycle and the provisional SWD average diurnal cycle. By calculating a diurnal amplitude index (DAI) from the original NWP filtered hourly time-series and we added in for each day an amplitude modulated diurnal signal (λ) to the SWD where

$$\lambda = \mathrm{DAI}\,(O(t) - \mathrm{SWD}_{\text{provisional}}(t)) \qquad (3)$$

where t=1,…,144 are the 10 minute time intervals within a day, O is the observed diurnal cycle, M is the diurnal cycle of the provisional SWD. The DAI was calculated by

$$\mathrm{DAI} = 2\,(W_{\text{early pm max}}) - (W_{\text{am min}} + W_{\text{late\_pm\_min}}) \qquad (4)$$

Note, an upper limit on the DAI was imposed to avoid “large” changes to the wind speed values

Step 8: Finally, the step which decomposed the residual (ε) in (Eq. 1) into a low-frequency (daily) and high-frequency (ten-minute) components resulted in power spectra with slightly less power at intermediate frequencies between about 4 and 16 hours. The coefficients at these frequencies were modified (consistently between sites) in such a way as to match better the behaviour of the observed spectra.

3. Results

In this section a sample of results is presented whereby properties of the observed site winds are compared against the MM5 based 5 year SWD and the NZLAM based 2 year SWD. Generally plots are presented for one or two sites only to indicate whether observed properties are matched by the SWD. However, plots have been generated for all sites and are available in pdf format upon request from the Commission. In Sections 3.1 - 3.4 ten-minute and hourly histograms of both wind speeds and deltas are presented for a selection of "better" and more "poorly" simulated sites. The use of the words "better" and "poorly" are relative and not intended as absolute judgements. In our judgement, histograms of wind-speeds and high frequency fluctuations at the 15 different sites have been reproduced very well for eleven of the fifteen sites, and reasonably well for the other four sites (CNI2, MWT1, CTY1, and STH1).

Plots for the entire periods of the SWD are shown as a sanity check on the entire SWD, while plots comparing just the period of overlap are shown to highlight differences with the observed and provide an indication if a particular period was unusual. The histograms of ten-minute deltas (Figure 7 in Section 3.2) do show a tendency for an under-prediction of the frequency of the -0.25 m/s to 0.25 m/s delta class. For 8 of the sites these are under-estimated by at least 5%, this means that for these sites the situations where the ten-minute winds are relatively light are under-simulated. These 8 sites are CTY1, CKS2, CKS3 (Figure 7a), MWT1, MWT3, CNI1, NTH1, and NTH2 (Figure 7b). This behaviour was not as apparent for the hourly deltas and likely resulted from the need to inflate the high-speed tail of the NWP histograms.

3.1 Wind speed histograms – ten minute

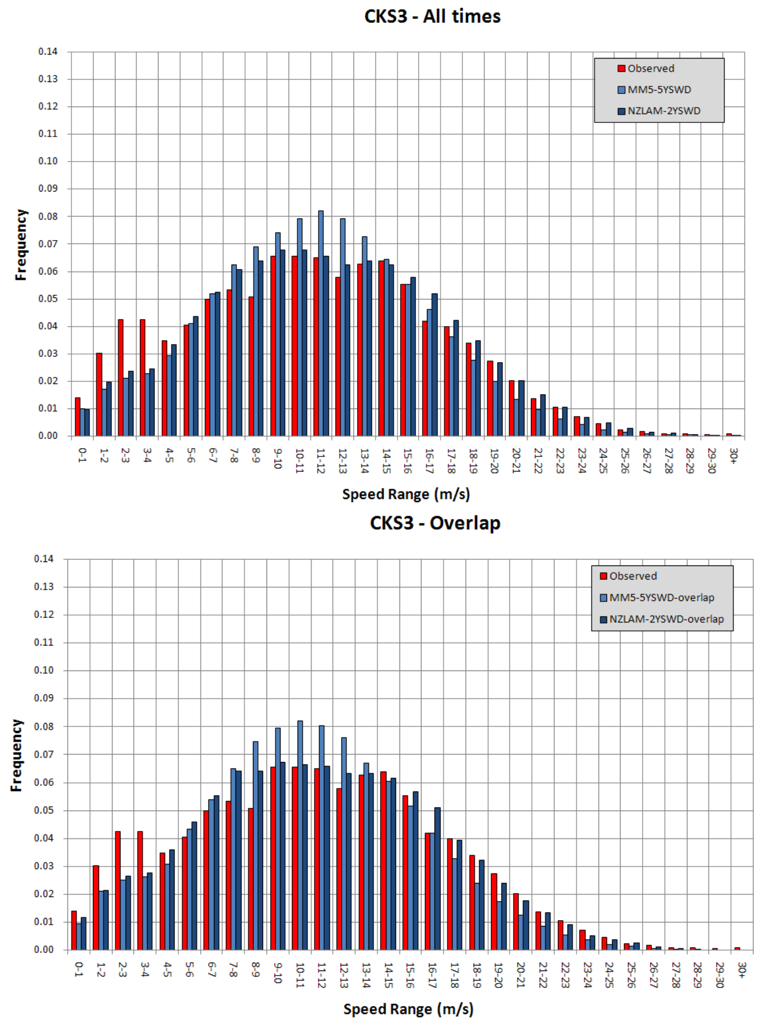


Figure 5: Histograms of observed, MM5-based SWD, and NZLAM-based SWD ten-minute wind speeds at site CKS3 for a) the entire periods for which each dataset is available and b) the period where the SWD overlaps with the 1 year period of observations.

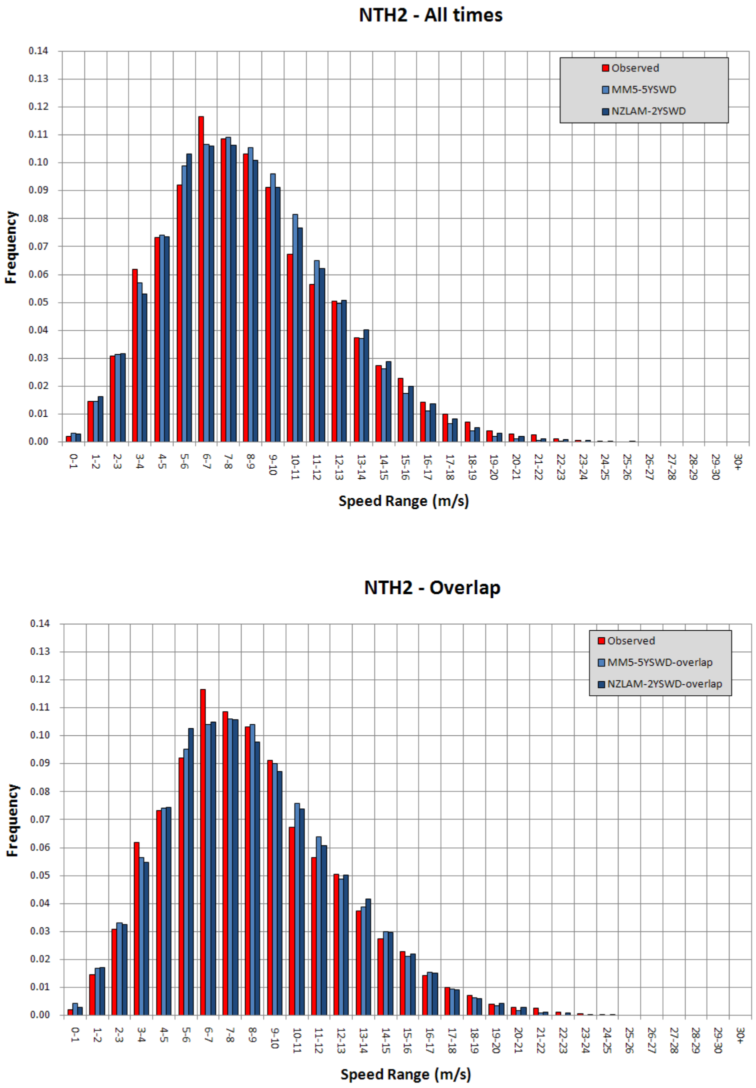


Figure 6: **Histograms of observed, MM5-based SWD, and NZLAM-based SWD ten-minute wind speeds at site NTH2 for a) the entire periods for which each dataset is available and b) the period where the SWD overlaps with the 1 year period of observations.**

3.2 Delta histograms – ten minute

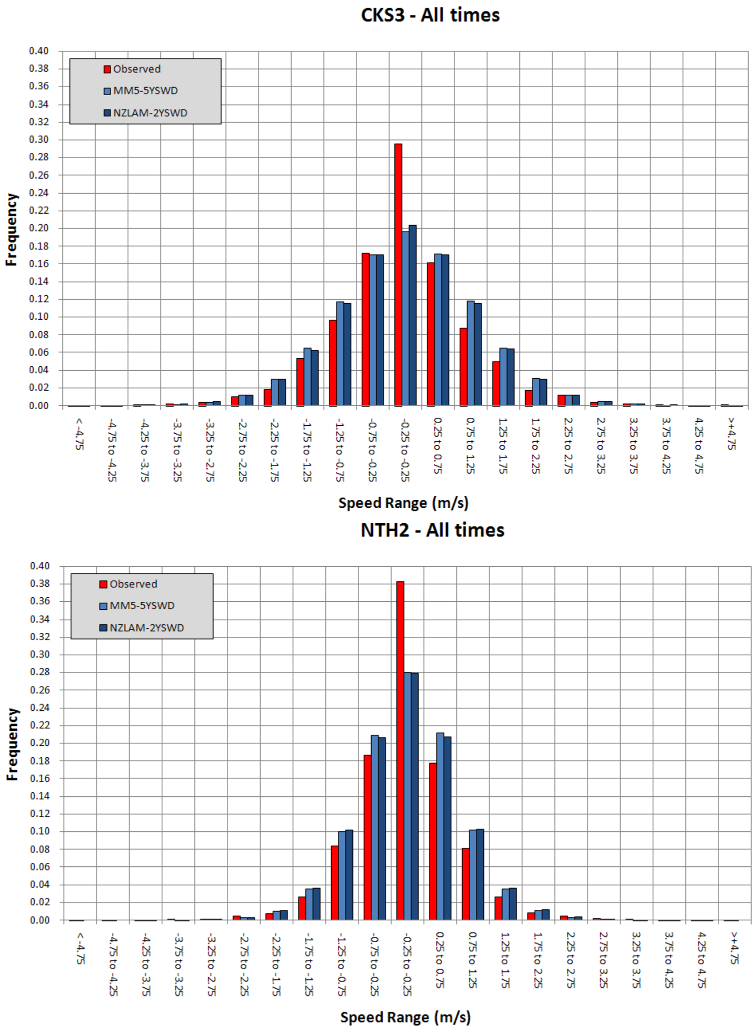
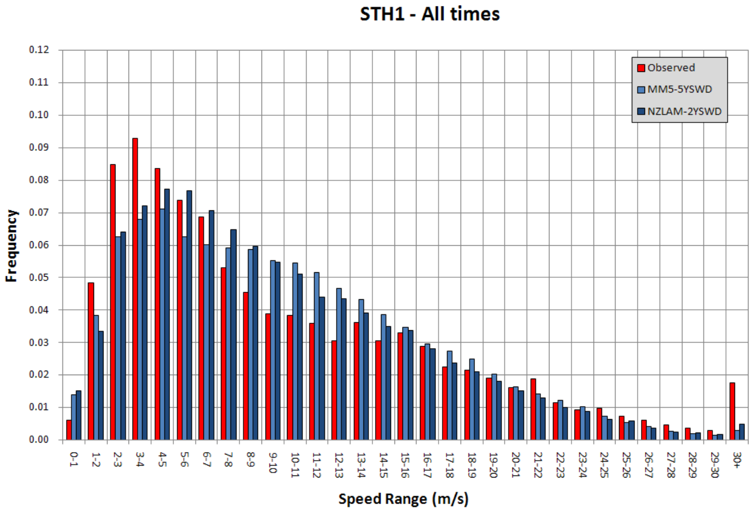


Figure 7: Histograms of observed, MM5-based SWD, and NZLAM-based SWD ten-minute wind speeds deltas at sites for the entire periods for which each dataset is available for sites a) CKS3 and b) NTH2.

3.3 Wind speed histograms – hourly



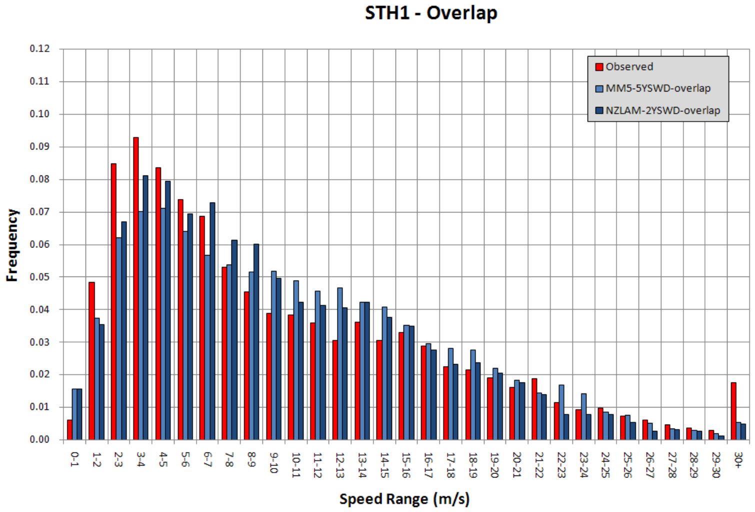


Figure 8: **Histograms of observed, MM5-based SWD, and NZLAM-based SWD hourly wind speeds at site STH1 for a) the entire periods for which each dataset is available and b) the period where the SWD overlaps with the 1 year period of observations.**

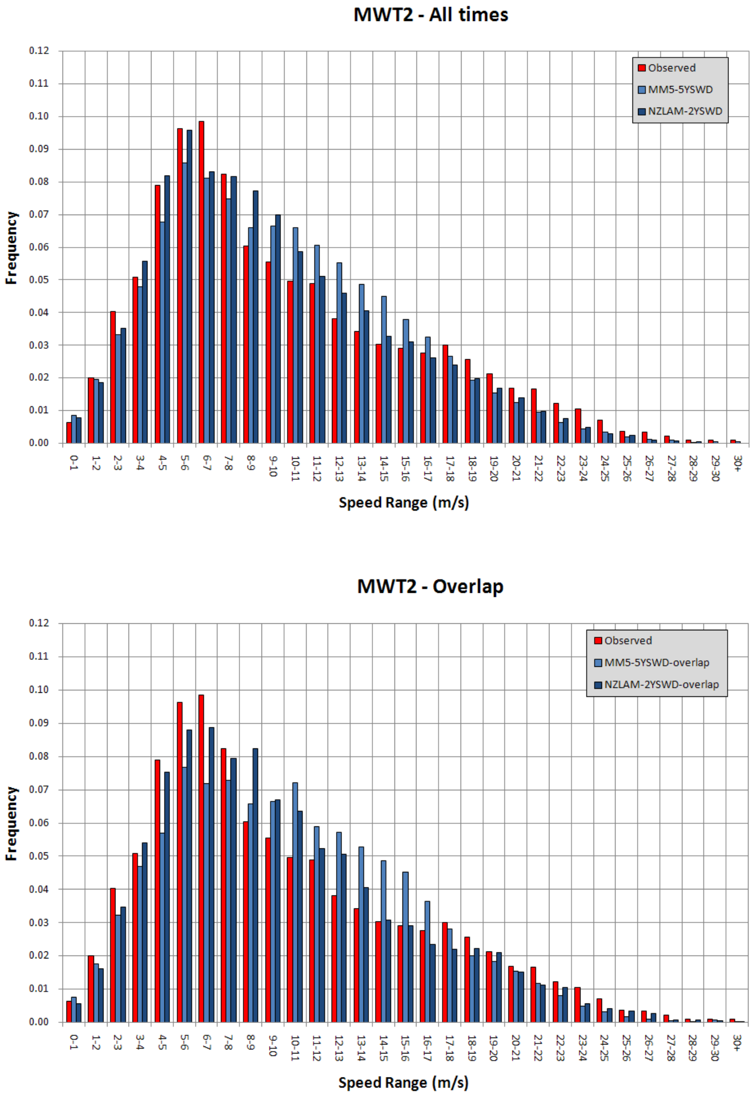


Figure 9: Histograms of observed, MM5-based SWD, and NZLAM-based SWD hourly wind speeds at site MWT2 for a) the entire periods for which each dataset is available and b) the period where the SWD overlaps with the 1 year period of observations.

3.4 Delta histograms – hourly

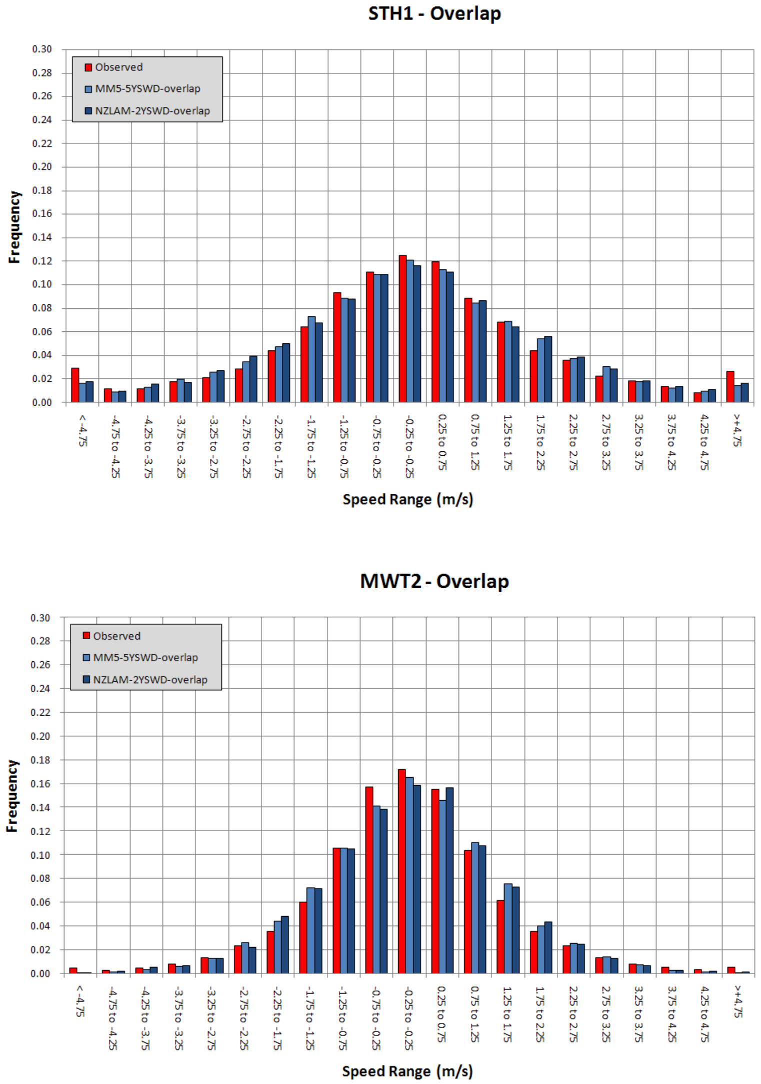


Figure 10: **Histograms of observed, MM5-based SWD, and NZLAM-based SWD hourly wind speeds deltas at sites a) STH1 for the entire periods for which each dataset is available and b) MWT2 for the period where the SWD overlaps with the 1 year period of observations.**

3.5 Monthly averages

Figure 11 shows that monthly averages were generally well captured in the SWD, indicating our method will probably provide good seasonal and inter-annual generating scenarios for the period of the SWD. In fact, it does indicate that if extended periods of NWP output, such as with a high resolution Regional Climate Model, were available that this kind of method could provide very good multi-decadal time-series.

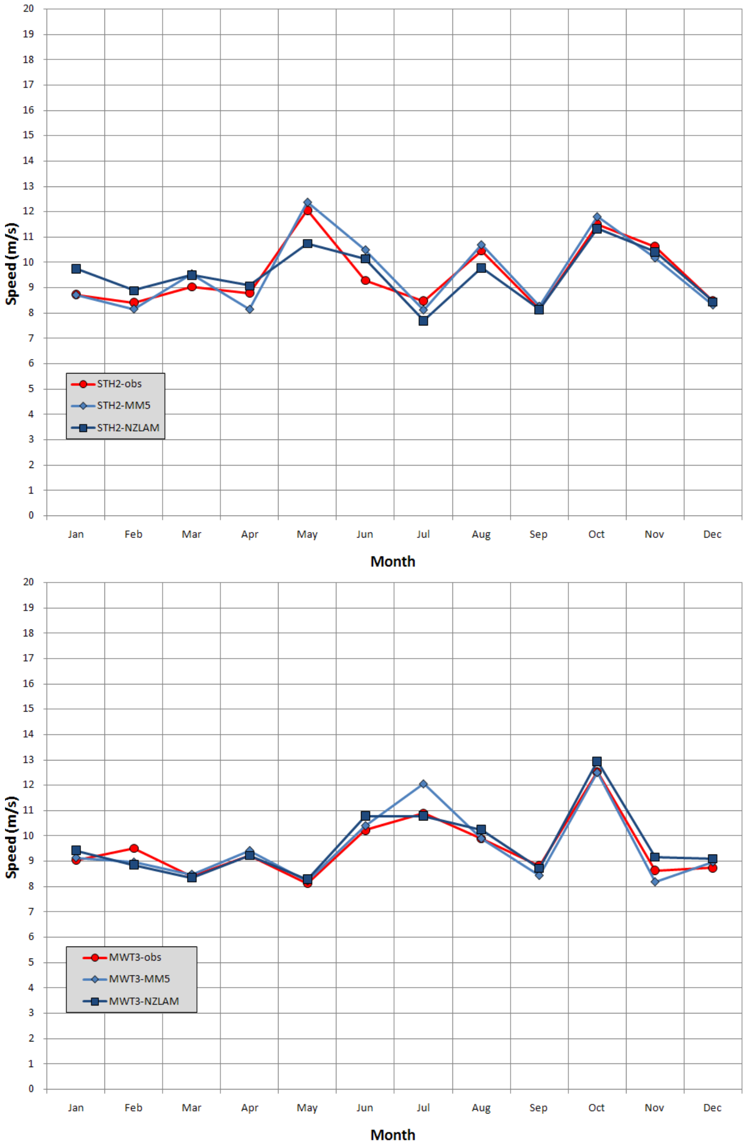


Figure 11: **Observed, MM5-based SWD, and NZLAM-based SWD monthly wind speeds for the period where the SWD overlaps with the 1 year period of observations at site a) STH2 and b) MWT3.**

3.6 Diurnal variations

Given that the methodology was done in such a way as to match the SWD diurnal cycle to the observed, the following plots (Figure 12) are presented to demonstrate the methodology worked as designed

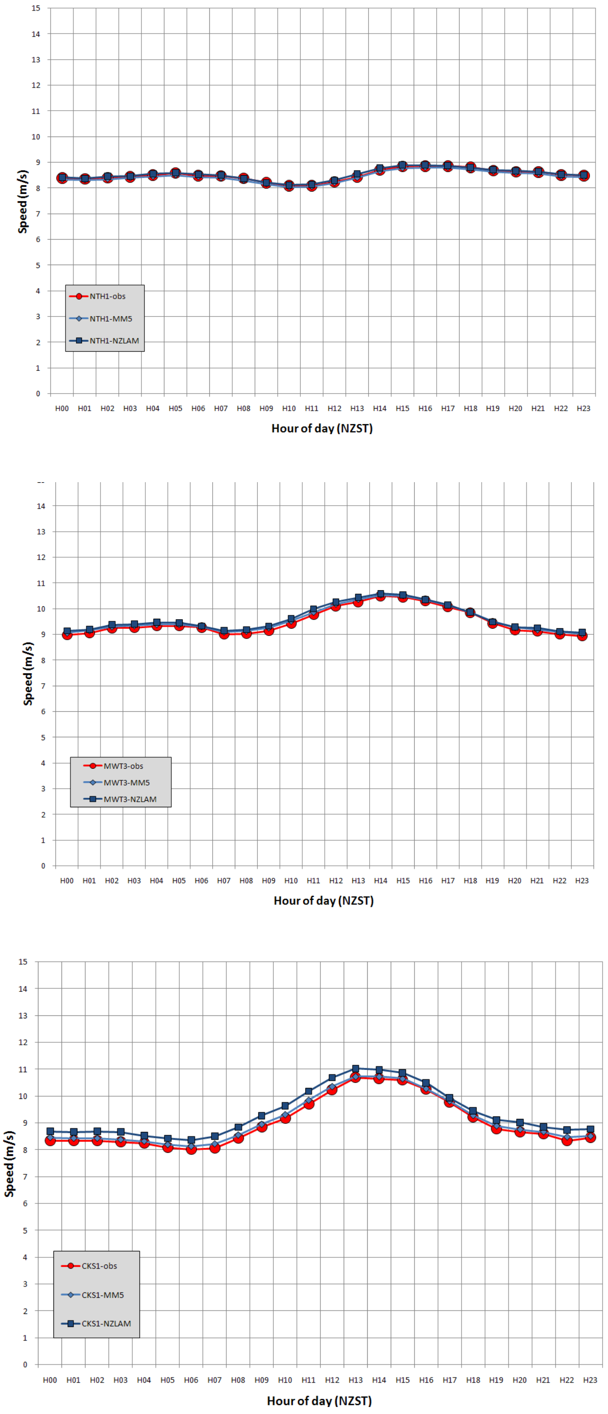


Figure 12: Observed, MM5-based SWD, and NZLAM-based SWD average diurnally variation of wind speeds for the entire periods for which each dataset is available at site a) NTH2, b) MWT3, and c) CKS1.

3.7 Power Spectra – 10 minute series (CKS1)

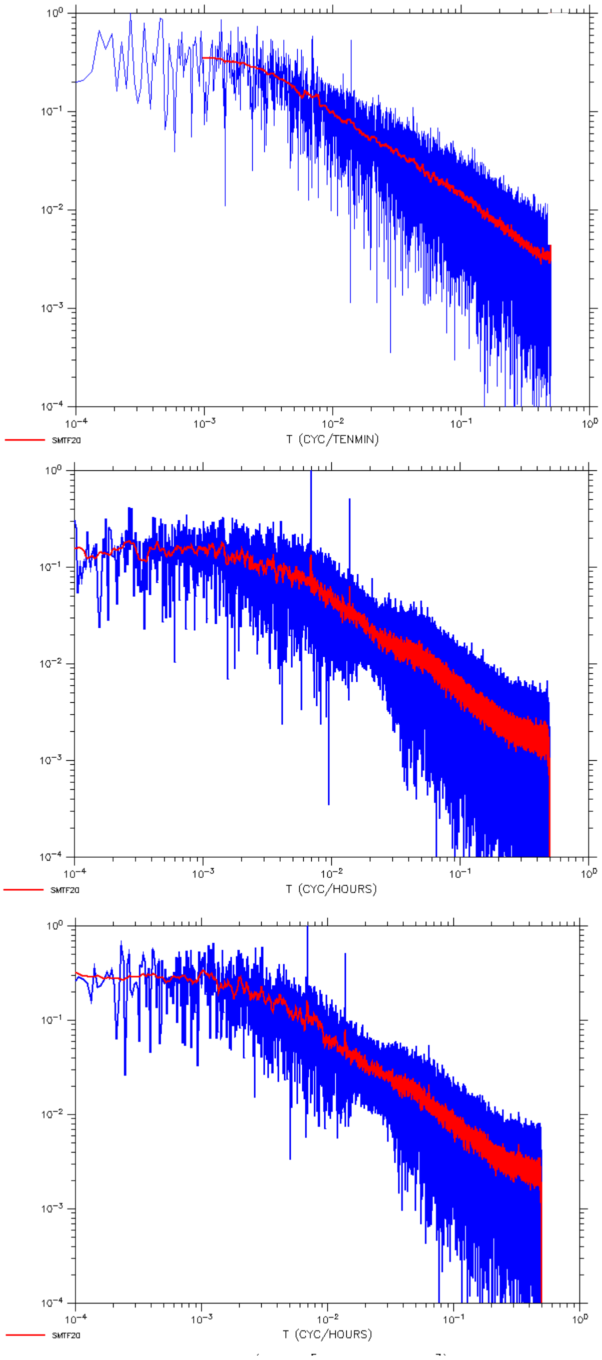


Figure 13: **Power spectra of wind speed times-series for the entire periods for which each dataset is available at site CKS1 for a) as observed, b) the MM5-based SWD, and c) the NZLAM-based SWD. The red curve is the highly smoothed moving average. (Note, these spectra are all for different periods.)**

3.8 Inter-site correlations of wind speeds

Observed inter-site correlations were calculated for each station pair where possible. Typically there were at least 30 days of overlap for the site pairs, with about half the sites having more than 180 days of overlap. However, there were two exceptions; CKS1 which had no overlap with any other site, and NTH3 where such a calculation would be not valid as the “observations” there were a dummy composite of two exposed “nearby” (within 100 km) climate stations. Figure 14 shows the inter-site correlations as a function of the distance between sites for both the original NWP hourly series and the provisional SWD’s (i.e., before step 5). Figure 15 shows the positive impact of step 5. The observed relationships seem to be generally preserved for both final SWD. The “minima” in correlations for distances of around 700 km is a “Cook Strait” effect, and is an artefact of the particular selection of 15 sites. For the interested reader a spreadsheet (inter-site-correlations.xls) with the actual numbers of overlapping day, distances between stations, and correlations is available along with the pdf’s.

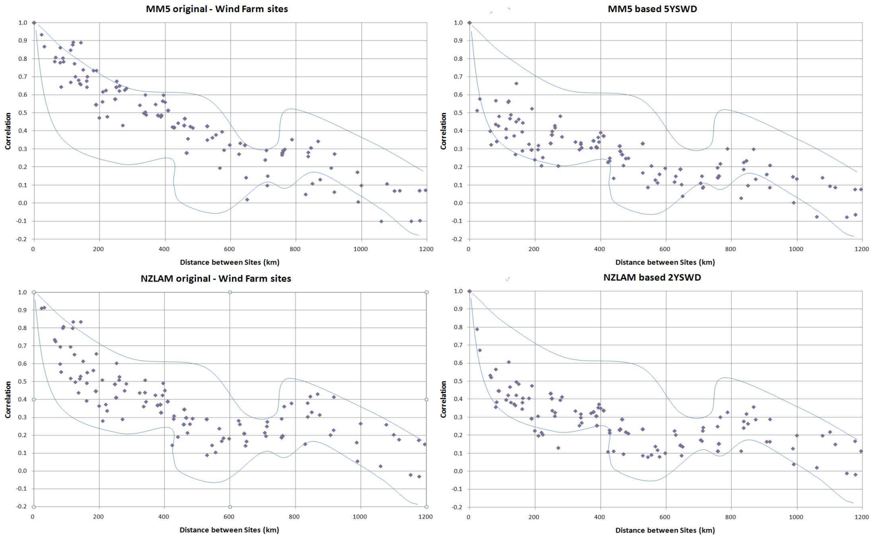


Figure 14: **Correlations of wind speeds between each of the 15 sites as a function of distance for a) the filtered hourly MM5 time-series (i.e., after step 2), b) the version of the MM5-based SWD after step 4 but before the inter-site adjustment, c) the filtered hourly NZLAM time-series (i.e., after step 2), and d) the provisional version of the NZLAM-based SWD after step 4 but before the inter-site adjustment The blue curves approximate the lower and upper bounds for the observed correlations.**

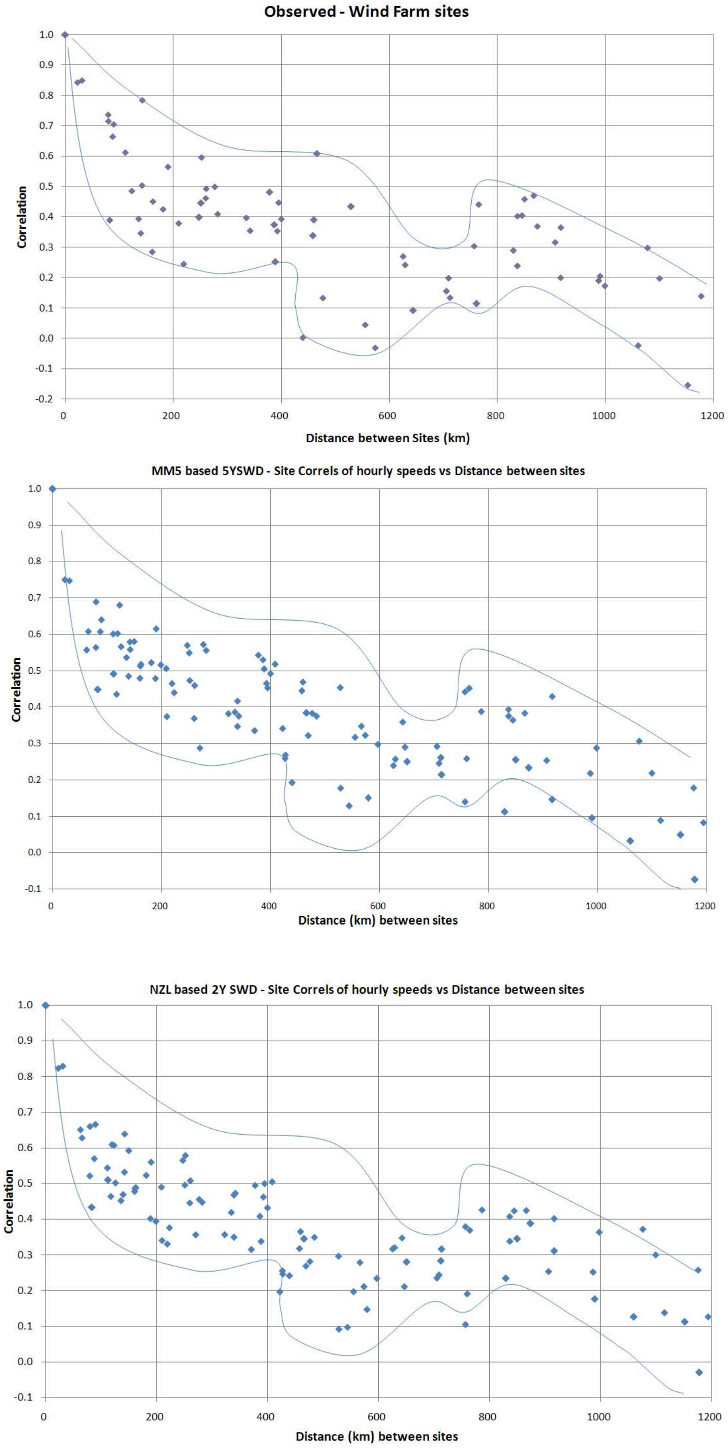


Figure 15: **Correlations of wind speeds between each of the 15 sites as a function of distance a) as observed (when possible), b) the MM5-based SWD, and c) the NZLAM-based SWD. The blue curves approximate the lower and upper bounds for the observed correlations.**

3.9 Inter-site correlations of deltas

Correlations of hourly and ten-minute deltas are shown in Figure's 16 and 17 respectively.

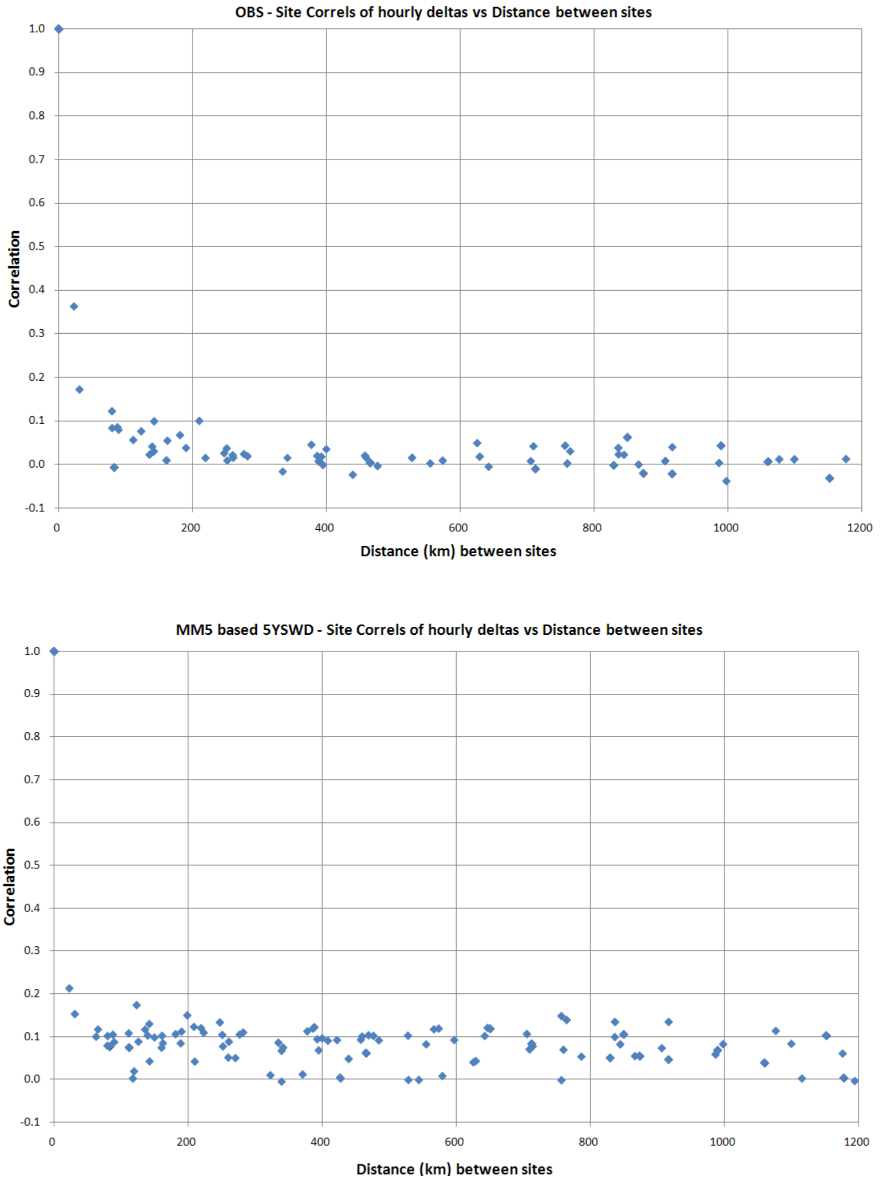
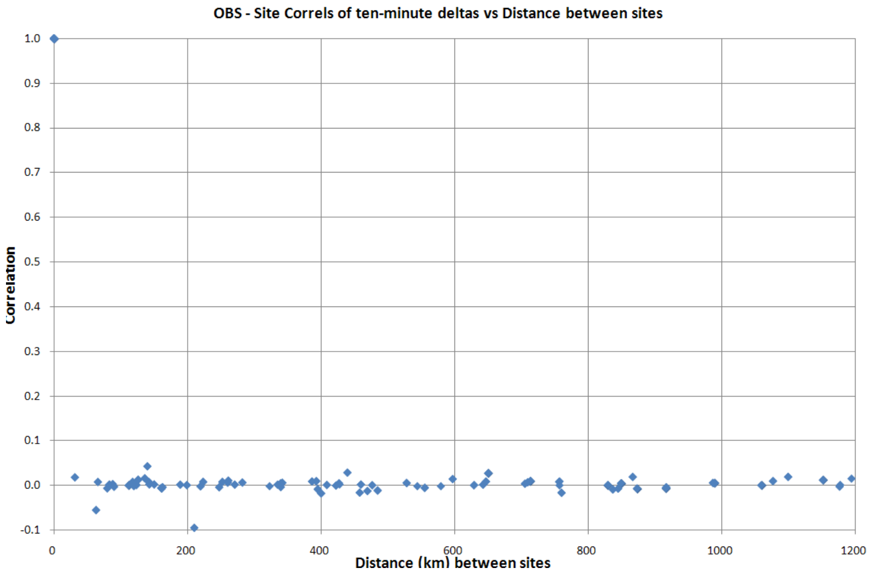


Figure 16: **Correlations of hourly deltas between each of the 15 sites as a function of distance a) as observed (when possible), b) the MM5-based SWD.**



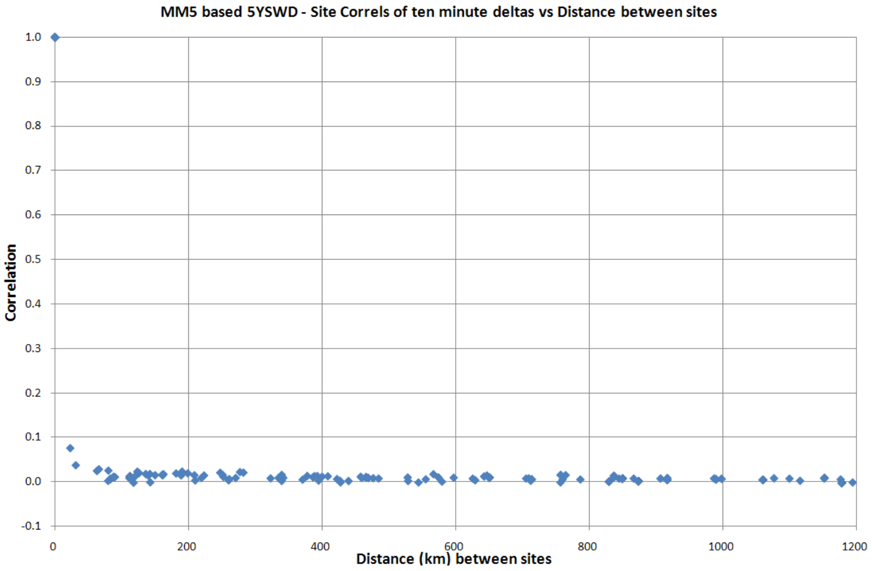


Figure 17: **Correlations of ten-minute deltas between each of the 15 sites as a function of distance a) as observed (when possible), b) the MM5-based SWD**

3.10 Time series for sub-periods

Figure 18 shows a segment of observed and SWD ten-minute time series for a 17 day period in June 2007 for the MWT1 and MWT2 sites. It shows how the relationships between the sites have been preserved in both SWD, but also how the large synoptically driven regional speed-up on the 4th and the decrease on the 9th have been captured by the SWD.

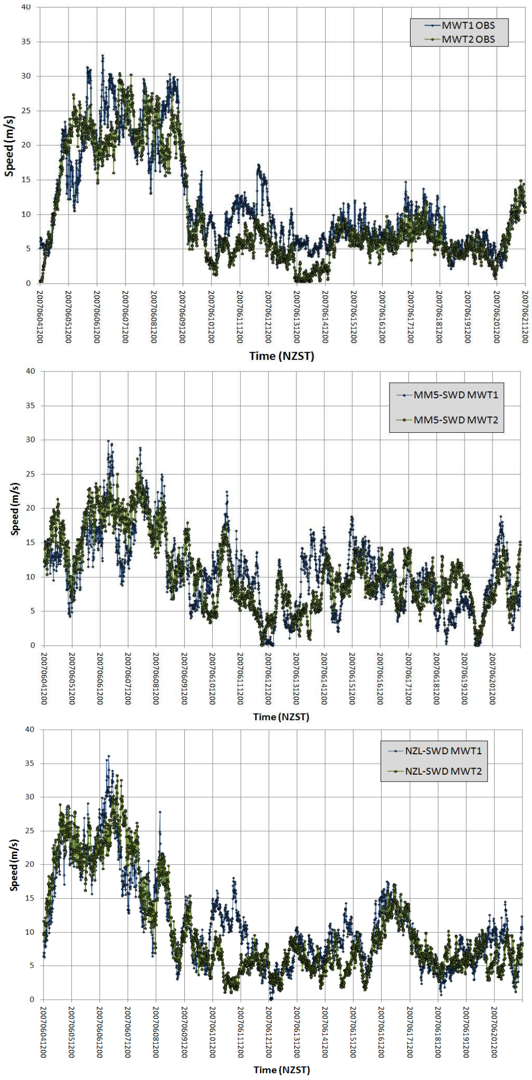


Figure 18: **Time series of ten-minute wind speeds at sites MWT1 and MWT2 for the period 1200 NZST 4 June to 1200 NZST 21 June 2007 as a) observed, b) from the MM5-based SWD, and c) from the NZLAM-based SWD.**

4. Conclusion and discussion

A 5 year (2 Sep 2003 to 30 Aug 2008) MM5-based 10-minute SWD and a 2 year (24 Oct 2006 to 15 Dec 2008) NZLAM-based 10-minute SWD at 85 metres altitude have been created for use in modelling the impacts of 15 actual or proposed wind farms on New Zealand's electricity system.

Summary statistics showing how well the two SWD correlate with observations for each site are given in Table 5. It seems that both SWD do a useful job of explaining the variance for all sites. Lower correlations at some of the sites (e.g., CKS1, CTY1, and NTH1) did not necessarily correspond to poorer representations of the histograms of wind speeds and delta's at those sites. In fact two sites, with quite good correlations STH1 and MWT1 had histograms with some differences in the 9-14 m/s ranges, tending to under-predict the frequency of winds in these ranges. Reassuringly, these two sites were able to reproduce quite well the power-generation curves calculated from the observed wind data, see Figure 19 which shows the curves for MWT1.

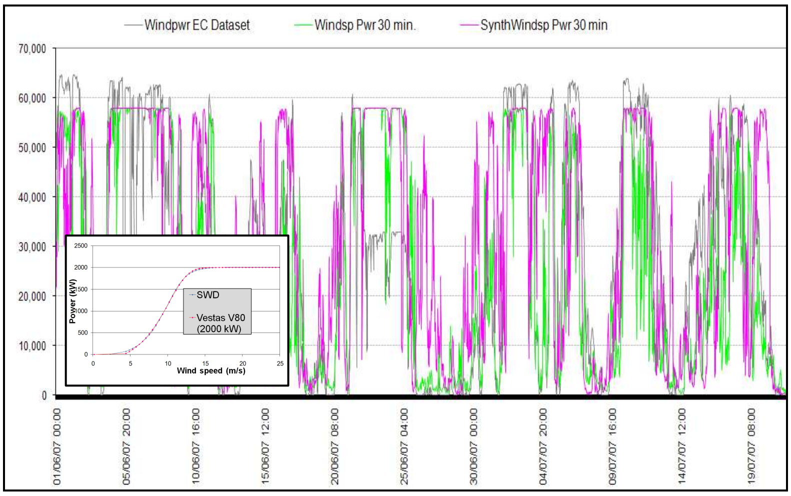


Figure 19. **Power output for a SWD site as estimated using observed wind speeds (grey curve) and the MM5-based SWD (green curve), as well as the actual power output (purple curve). The major grid interval on the x-axis is 1 day. The inset shows that there is little discernible difference between the turbine "PowerCurves".**

Differences between the two SWD are likely due to (i) NZLAM employing a 3DVAR data assimilation scheme whereas MM5 does not and so has a longer spin-up period, (ii) the MM5 SWD was based on selecting output from a single suitable nearby grid point whereas for NZLAM interpolated values to the general location of the site were used, and (iii) for the MM5 SWD wind speeds from the filtered hourly series were interpolated between 75 m and 100 m levels to 85 m, while for NZLAM power-law extrapolation of 10 m winds to 85 m was done.

Overall, the demands required of the SWD have been met and the realistic preservation of the important properties of the available short term ten minute mast data, albeit disguised, has been demonstrated here.

Table 5: **Bias and explained variance of observed wind speeds by original filtered MM5 and NZLAM hourly winds for each site as well as the explained variance of the final SWD. Note, for CKS1 different sets of observations were used. Biases afterwards were generally close to 1.**

Site	MM5 Bias Before	MM5 r^2 Before	MM5-SWD r^2 After	NZLAM Bias Before	NZLAM r^2 Before	NZL SWD r^2 After
STH1	0.99	0.64	0.708	0.44	0.60	0.690
STH2	1.00	0.58	0.657	0.53	0.66	0.712
STH3	1.04	0.53	0.660	0.49	0.60	0.720
CTY1	0.94	0.45	0.592	0.68	0.48	0.552
CKS1	0.90	0.41	0.480	0.76	0.68	0.540
CKS2	0.89	0.53	0.637	0.88	0.70	0.778
CKS3	0.76	0.33	0.519	0.65	0.61	0.699
MWT1	0.84	0.49	0.596	0.49	0.64	0.688
MWT2	1.00	0.61	0.662	0.57	0.68	0.725
MWT3	0.93	0.51	0.557	0.62	0.59	0.628
CNI1	0.91	0.70	0.770	0.45	0.58	0.690
CNI2	0.99	0.72	0.592	0.46	0.62	0.532
NTH1	0.91	0.62	0.490	0.73	0.68	0.486
NTH2	0.99	0.67	0.704	0.92	0.72	0.766
NTH3	1.00	0.38	0.590	0.92	0.39	0.595

5. Acknowledgements

We would like to thank Brian Bull (Electricity Commission) and Devin Kilminster (MetService) for input into statistical methods during the course of the project. We also appreciate very much the following companies (contacts in brackets) for providing the observed ten minute mast data used in the study: Meridian (Paul Botha), Contact (David Rohan), TrustPower (Clayton Delmarter), Unison(Daniel Stettner), and MainPower (Andrew Hurley).

6. References

Bloom, S.C., L.L. Takaks, A.M. Da Silva and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Monthly Weather Review*, **124**, 1256 – 1271.

Davies, T., Cullen, M.J.P., Malcolm, A.J., Mawson, M.H., Staniforth, A., White, A.A., Wood, N. A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, **131**, 1759 – 1782, 2005.

Grell, G., A., J. Dudhia, and D. R. Stauffer, 1994: A Description of the fifth Generation Penn State/NCAR Mesoscale Model (MM5). NCAR Technical Note TN-398+IA, NCAR 117 pp. *(http://www.mmm.ucar.edu/mm5/documents/mm5-desc-doc.html)*.

Haslett, J. and A. E. Raftery, 1989: Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Appl. Statistics*, **38**, No 1. 1-50.

Lorenc, A.C., Ballard, S.P., Bell, R.S., Ingleby, N.B., Andrews, P.L.F., Barker, D.M., Bray, J.R., Clayton, A.M., Dalby, T., Li, D., Payne, T.J., Saunders, F.W., The Met. Office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, **126**, 2991 – 3012, 2000.

Reid S., and R. Turner, 2001: Correlation of real and model wind speeds in complex terrain. *Weather and Forecasting*, **16**, 620-627.

Staniforth, A., and multitude of co-authors, (2004). Joy of U.M. 6.0. Model formulation. Unified Model Documentation paper No. 15, version 6.0. Technical note from The MetOffice™ (United Kingdom).

Webster, S., M. Uddstrom, H. Oliver, and S. Vosper, 2008: A High Resolution Modeling Case study of a severe weather event over New Zealand. *Atmospheric Science Letters*. **9(3)**:119-128. DOI 10.1002/asl.172.

7. Appendix A: Survey of a “Haslett and Raftery” (non-NWP) type approach

The common purpose of both pieces of work i.e., Haslett and Raftery (1989) (HR89) and the current EC-SWD project is the generation of long period wind series at sites with observations available only for a short period.

The basic analysis steps of HR89 are:

- Estimation of signal
- Estimation of residual
- Estimation of spatial dependence

These are also the steps we followed in our analysis. However, due to the differences between the datasets being analysed, these three steps were modified to improve our analysis. A comparison of the major features of HR89 and EC-SWD is summarised in Table A1.

For the current project, applying Eq. (A1) used by HR89 significantly reduced the explained variance

$$w = a + bx + \sum_{i=-3}^{1} c_i y_i + d_1 \,|\cos(z)| + d_2 \,|\cos(z + 45^o)| + \varepsilon \qquad \text{(A1)}$$

To investigate the performance of Step 1 in HR89, we performed a case study for station MWT1. Here, nearby surrounding climate stations with long ten-minute records were Dannevirke and Palmerston North.

The explained variance obtained for the HR89 method at MWT1 was 0.520, while for our approach it was 0.596 for the MM5-based SWD and 0.688 for the NZLAM based SWD. This shows that the current approach was comparable to or better than HR 1989. Given that all other sites for this project don’t have such an ideal situation of having two close nearby stations, we considered it not worthwhile to pursue the HR89 method. This is especially so given the need to fill in sensibly, i.e., synoptically consistent winds for periods of missing data that do exist in the long-term climate station observations.

Table A1. Data and analysis method for the HR89 and the current EC SWD project.

	HR 1989	EC SWD project
Data with long period	11 stations over Ireland, with observations	MM5 or NZLAM model forecast hourly winds at the sites
Data format	Square root of daily wind average, no wind direction data is recorded.	Observation: 10 minute speeds. Modelled: hourly speeds plus direction.
1) Signal	Estimated by regression with the observation with long record at neighbouring stations.	See Eq. (A1) for the more details.
2) Residual model	arima(p,d,q), d<0.5, daily time scale	AR(6) with 10 minute time scale.
3) Spatial dependence	Statistical estimated (Kriging)	Large part of daily and hourly spatial dependence are accounted for by MM5 or NZLAM. Statistical estimation can be applied for the spatial dependence of the residuals.

8. Appendix B: Regression models (coefficients) for each site.

Table A2: Coefficients for Eq. 1 and A1 as developed for MM5.

Site	a	b	c_0	c_{-1}	c_{-2}	c_{-3}	c_1	d_1	d_2
STH1	-1.734	1.055	0.141	0.116	0.065	0.291	0.169	0.728	1.25
STH2	3.565	0.937	0.121	0.141	0.074	0.185	0.166	-1.391	-3.747
STH3	2.331	0.968	0.08	0.119	-0.007	0.255	0.135	0.586	-4.466
CTY1	-1.199	0.893	0.04	0.084	0.061	0.175	0.151	0.44	3.467
CKS1	1.047	0.771	0.074	0.045	0.068	0.035	0.238	-1.175	3.141
CKS2	-3.353	1.018	0.073	0.052	0.042	0.108	0.33	0.815	5.2
CKS3	-2.327	0.843	0.105	0.147	-0.022	0.238	0.226	3.563	4.65
MWT1	1.463	0.995	0.062	0.112	0.069	0.266	0.182	-3.411	2.592
MWT2	-0.065	1.019	0.072	0.046	0.009	0.199	0.255	-3.114	2.175
MWT3	3.114	0.727	0.078	0.098	0.102	0.186	0.163	-0.215	0.139
CNI1	-1.37	0.979	0.055	0.125	0.059	0.144	0.199	0.286	2.785
CNI2	3.375	0.678	0.036	-0.044	-0.057	-0.066	-0.313	-2.153	-0.549
NTH1	3.846	0.783	0.017	-0.056	-0.099	-0.203	-0.264	-0.933	-2.676
NTH2	2.415	0.797	0.074	0.066	0.072	0.219	0.162	0.342	-1.567
NTH3	2.387	0.782	0.096	0.081	0.045	0.285	0.073	-1.213	0.462

Table A3: Coefficients for Eq. 1 and A1 as developed for NZLAM

Site	a	b	c_0	c_{-1}	c_{-2}	c_{-3}	c_1	d_1	d_2
STH1	-2.154	2.158	0.232	0.2	0.274	0.399	0.309	3.721	-0.526
STH2	2.388	1.526	0.196	0.286	0.16	0.32	0.281	0.325	-1.042
STH3	0.9	2.116	0.074	0.363	0.323	0.461	0.342	-3.64	2.292
CTY1	-3.55	1.341	0.104	0.196	0.15	0.215	0.168	4.735	2.285
CKS1	-1.221	1.084	0.029	0.145	0.179	0.29	0.128	2.944	0.728
CKS2	0.17	1.062	0.143	0.286	0.131	0.212	0.07	0.372	0.327
CKS3	-2.343	1.206	0.211	0.102	0.157	0.229	0.311	5.523	0.275
MWT1	-0.206	1.75	0.223	0.315	0.203	0.278	0.337	0.828	1.352
MWT2	-1.323	1.496	0.004	0.159	0.201	0.108	0.421	0.736	3.134
MWT3	4.076	1.245	0.136	0.202	0.173	0.297	0.277	-1.296	-1.324
CNI1	-2.268	2.177	0.063	0.084	0.111	0.283	0.268	2.036	1.53
CNI2	1.509	1.427	0.094	-0.1	-0.043	-0.12	-0.482	0.13	0.741
NTH1	2.746	1.122	0.058	-0.079	-0.013	-0.27	-0.46	-1.22	-0.869
NTH2	2.341	0.91	0.11	0.097	0.075	0.211	0.145	-0.763	-0.951
NTH3	1.069	0.968	-0.004	0.166	0.019	0.305	0.119	-0.902	1.131

Table A4: Autocorrelation coefficients for MM5 SWD model

Site	α_1	α_2	α_3	α_4	α_5	α_6	σ^2
STH1	0.945	-0.065	0.031	0.005	0.019	0.016	1.914
STH2	0.943	-0.072	0.051	0.014	0.011	0.009	0.971
STH3	1.016	-0.121	0.041	-0.003	0.005	0.018	1.718
CTY1	0.998	-0.087	0.019	0.01	-0.001	0.022	0.804
CKS1	1.065	-0.163	0.031	0.009	-0.009	0.026	1.027
CKS2	0.905	-0.016	0.029	0.019	0.004	0.022	0.887
CKS3	0.907	-0.034	0.037	0.009	0.005	0.029	1.114
MWT1	0.929	-0.041	0.057	0.009	-0.001	0.015	1.05
MWT2	0.879	0.01	0.027	0.026	0.005	0.022	0.911
MWT3	0.891	-0.023	0.048	0.016	0.001	0.023	0.91
CNI1	0.972	-0.074	0.026	0.013	0.002	0.019	0.567
CNI2	0.946	-0.064	0.038	0.026	-0.003	0.029	0.658
NTH1	0.926	-0.075	0.057	0.034	0.005	0.029	0.692
NTH2	0.822	0.007	0.048	0.026	0.015	0.03	0.568
NTH3	0.977	0	0	0	0	-0.023	0.657

Table A5: Autocorrelation coefficients for NZLAM model

Site	α_1	α_2	α_3	α_4	α_5	α_6	σ^2
STH1	0.932	-0.054	0.023	0.015	0.008	0.028	1.857
STH2	0.932	-0.073	0.049	0.017	0.011	0.015	0.929
STH3	1.003	-0.12	0.039	-0.002	0.002	0.021	1.731
CTY1	0.996	-0.084	0.017	0.011	-0.002	0.022	0.856
CKS1	0.981	0	0	0	0	-0.02	0.844
CKS2	0.853	-0.006	0.045	0.012	0.012	0.023	0.827
CKS3	0.907	-0.034	0.037	0.009	0.005	0.029	1.114
MWT1	0.917	-0.045	0.062	0.006	0.002	0.019	0.986
MWT2	0.867	0.014	0.032	0.016	0.003	0.031	0.868
MWT3	0.882	-0.025	0.043	0.018	0.002	0.024	0.921
CNI1	0.979	-0.073	0.026	0.014	0.004	0.019	0.561
CNI2	0.948	-0.069	0.041	0.026	-0.001	0.03	0.638
NTH1	0.921	-0.073	0.058	0.034	0.003	0.031	0.681
NTH2	0.812	0.009	0.047	0.027	0.012	0.033	0.562
NTH3	0.977	0	0	0	0	-0.023	0.655

9. Appendix C: Format of SWD files

The ten-minute SWD files are .csv files and are named 2ySWD-EC-tenm.csv and 5ySWD-EC-tenm.csv. The 2ySWD prefix indicates a NZLAM-based SWD, while the 5ySWD indicates a MM5-based SWD. Hourly average files are also available.

The columns of the csv file are given in the header row. These are

Year,Month,Day,Hour,Minute, Speed (m/s) for Site 1, …, Speed (m/s) for Site 15

Where

Site 1 = STH1
Site 2 = STH2
Site 3 = STH3

Site 4 = CTY1
Site 5 = CKS1
Site 6 = CKS2

Site 7 = CKS3
Site 8 = MWT1
Site 9 = MWT2

Site 10 = MWT3
Site 11 = CNI1
Site 12 = CNI2

Site 13 = NTH1
Site 14 = NTH2
Site 15 = NTH3