

Appendix A

The GRAMM-GRAL system

A1 Model development and description

A1.1 Background

The development of GRAL began at the Graz University of Technology (TUG) in 1999. Since 2006 the further development of the model has involved a co-operation between TUG and the Government of Styria (Öttl, 2016b). GRAL is now incorporated into the SoundPLAN software¹.

The initial driver for GRAL was the need for an accurate and practical model that could be applied to complex terrain and low-wind-speed conditions (< 1.5 m/s) in the inner-Alpine basins of Austria. Another important aspect of model development has been the simulation of the dispersion of pollutants from road tunnel portals. The features of GRAL have subsequently been extended, and the current version of the model also includes the following:

- Modelling on a wide range of spatial scales. The scale of application ranges from individual streets to whole cities that are tens of kilometres across, or even larger areas where topography and land use determine wind flows.
- Simultaneous modelling of multiple source types, including surface road networks, point sources such as tunnel ventilation outlets and industrial stacks, tunnel portals, and area sources.
- Incorporation of a full year of meteorological data.
- Dispersion over the full range of wind speeds without any lower threshold, and for all stability conditions. GRAL has the ability to predict concentrations for low-wind-speed conditions (< 1 m/s) better than most Gaussian models (e.g. CALINE).
- Allowance for the effects of buildings, including building downwash effects through microscale modelling.

GRAL has been used extensively in regulatory assessments and scientific studies, and fulfils the requirements of the Technical Guideline RVS 040212 for dispersion modelling in Austria. It is most frequently used to calculate the impacts of road traffic on air pollution, but it has also been used to assess other sectors (e.g. technical regulations for domestic heating). GRAL has been extensively validated in a number of different countries and contexts.

A1.2 Overview of model system

The most detailed descriptions of the GRAMM-GRAL model system are provided in the following documents:

- Documentation of the prognostic mesoscale model GRAMM (Graz Mesoscale Model) version 16.1 (Öttl, 2016a).
- Documentation of the Lagrangian Particle Model GRAL (Graz Lagrangian Model) version 16.8 (Öttl, 2016b).
- Recommendations when using the GRAL/GRAMM modelling system (Öttl, 2016c).
- GRAL Manual - GRAL Graphical User Interface 16.8 (Öttl and Kuntner, 2016).

¹ <http://www.soundplan.eu/english/soundplan-air-pollution/gral-system/>

The system was developed with the modelling of complex urban road networks and tunnel portals as core capabilities. An overview of the system is presented in Figure A-1, and the different elements are described in the following sections.

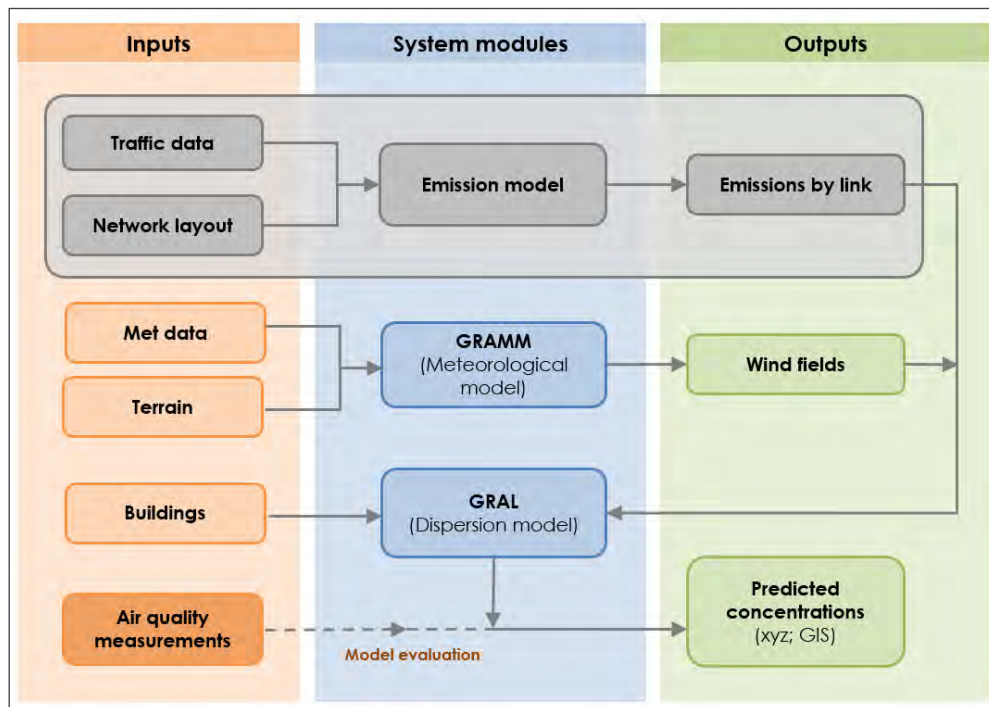


Figure A-1: Overview of the GRAMM-GRAL modelling system

A1.3 GRAMM

A1.3.1 General principles

The model system has three different meteorological pre-processing approaches, depending on the input data. It is possible to use an hourly time-series of meteorological data directly, as well as statistics of wind speed, direction, and stability class. In the case of complex terrain, GRAL can be coupled with GRAMM.

GRAMM is an Eulerian prognostic, mesoscale wind field model that is the meteorological driver for GRAL. GRAL is a Lagrangian dispersion model. Eulerian and Lagrangian models consider the motion of fluids (in this case parcels of air) through a field in different ways. In an Eulerian model the frame of reference is fixed, and the parcels of air pass through it. In a Lagrangian model the trajectories of the parcels of air are followed as they move through the field. The physics of GRAMM are complex and a full description is not included here; the reader is directed to the model documentation for a more in-depth description (Öttl, 2016a).

The main features of GRAMM include:

- A prognostic non-hydrostatic wind field model.
- A terrain-following grid. At all scales the effects of terrain on dispersion (e.g. cold air drainage flows) are taken into account.
- The computation of the surface energy balance, which accounts for topographic shadowing effects.

In contrast to a diagnostic wind field model, a prognostic model enables the representation of dynamic effects due to local topography such as obstacle-influenced air flows, and is capable of accommodating complex topography with high a horizontal resolution (Uhrner et al., 2014). A grid resolution of less than 10 metres is possible in GRAMM, although larger grid cells tend to be required for larger areas to maintain acceptable processing times. It is generally recommended to apply the Match-to-Observation and/or the Re-Order-function after GRAMM simulations.

Because the simulation of an hourly time series of concentrations for a whole year would be time-consuming, GRAMM computes steady-state three-dimensional wind fields for specific 'meteorological situations'. The meteorological situations are obtained from the one-hour time series of meteorological measurements at a point within the study area, and are defined based upon seven stability classes (function of wind speed and solar radiation), 36 wind direction classes, and several user-defined wind speed classes. Around 1,000 different meteorological situations are usually sufficient to characterise the dispersion conditions in an area during all 8,760 hours of the year.

The developers recommend that the GRAMM Re-Order and/or Match-to-Observation functions (see below) are applied after the initial simulations have been completed (Öttl D and Kuntner M, 2016).

A1.3.2 Re-Order function

At the end of GRAMM simulation a wind field is stored for each meteorological situation. The Re-Order function searches these wind fields to determine the best fit against the observed data at the location of the measurement site. In this way a significant improvement in the simulated wind data at the monitoring site used as input to GRAMM can be achieved. The more flow fields that have been computed, the better the results of the Re-Order function.

A1.3.3 Match-to-Observations function

This function aims to match existing the GRAMM wind fields to any meteorological observations inside a domain, regardless the period of time when these measurements have been taken. The imported time series of meteorological data is synchronised automatically. Thus, it is not necessary to have each time series covering exactly the same time period. The Match-to-Observation procedure opens up an additional modelling strategy with GRAMM. In a first step the simulations can be carried out using artificial data comprising all theoretical possible classified situations. In the second step these wind fields can be used to match any new meteorological observations inside the domain as long as no other adjustments are made. The more flow fields are available for the fitting process, the better the results of the Match-to-Observation function.

Where Match-to-Observations is used for multiple reference stations the result will be a compromise. The match is optimised across all stations, and therefore the overall model performance should improve. However, for any given station the predictions may or may not improve, particularly where the meteorological data across multiple stations in a domain are dissimilar (as in this study).

A1.4 GRAL

A1.4.1 General principles

As GRAL is a Lagrangian model, pollutant concentrations are predicted by simulating the movement of individual 'particles' of a pollutant emitted from a source along trajectories in a three-dimensional wind field. The trajectory of each of the particles is determined by a mean wind velocity component and a fluctuating (random) wind velocity component. In order to compute particle trajectories GRAL needs to know the velocity

components all over the computational domain. The particle concentration within any given grid cell is then computed as the density of the trajectory points. In each cell, and for each time step, the instantaneous particle concentration can be obtained by summing up the masses of the particles being located in the same cell at that time and dividing that number by the volume of the cell (Uhrner et al., 2014).

To obtain the wind velocity components, GRAL can use either (i) the flow fields computed by GRAMM, by interpolating the output of the latter over its own grid, or (ii) wind velocities calculated within GRAL using a simplified flow solver. The latter is summarised by Uhrner et al. (2014). GRAMM was developed as a prognostic model to predict meteorological conditions in mountainous areas where local winds are generated by mountain/valley winds, strong topography effects, etc. It was developed for regions where very dense network of meteorological measurements would be needed to run a diagnostic wind field model (e.g. CALMET). For quite flat regions and ground-based emission sources (e.g. roads) GRAL would normally not need to run GRAMM as the features of GRAL allow for consideration of all the local influences to the flow field close to the source.

GRAL includes specific algorithms to take into account meandering under low wind speed conditions, and dispersion from the portals of road tunnels.

General parameters required by GRAL include surface roughness length, dispersion time, number of traced particles (which influences the statistical accuracy of the results) and counting grids (variable in all three directions).

A1.4.2 Pollutants

GRAL only models the transport and diffusion of chemically non-reactive gases, with particles also being treated as an unreactive gas. Chemical reactions are not taken into account in the dispersion calculations, and for roads and diffuse sources this typically leads to a slight overestimation of concentrations. The formation of secondary organic and inorganic PM is also not considered.

A1.4.3 Emission sources/source groups

Multiple 'source groups' can be defined in GRAL to represent different emission sources. These include:

- Point sources (e.g. industrial stacks).
- Line sources (e.g. roads, railway lines). Line sources are defined as road bands with 3 metre vertical expansion (for vehicle turbulence).
- Area sources (e.g. domestic heating).
- Tunnel portals. The development of the tunnel portal module of GRAL is described in detail in Öttl et al. (2002). As the GRAL optimisation project will not deal with tunnel portals this module has not been described further.

In theory there is no limit to the number of separate emission sources that can be included in a GRAL simulation, although there is always a practical limit.

A1.4.4 Emissions

An annual average emission rate (in kg/hour) is defined for each source group in GRAL. For road traffic GRAL has in-built algorithms for calculating emission rates (the grey area of Figure A-1). These are based on the road traffic emission model NEMO, which was also developed at the Technical University of Graz. The input to NEMO consists of a road network, street classifications and other attributes, such as the

average road gradient and the annual average daily traffic. NEMO provides the emission rate for each road segment (in kg/km/h), based on emission factors and fleet models. In this study the NEMO emission factors were not used, but were replaced by emission rates developed by NSW EPA (see section 4.2).

A1.4.5 Emission modulation factors

Diurnal and seasonal variations in emissions from each source group can be characterised using 'modulation factors', or else a time series of emissions can be defined. The final result is a time series of concentrations that is dependent on the classified meteorological situations and the seasonal and diurnal modulation factors (or time series of emissions). In this way annual mean, percentile, maximum daily mean, or maximum concentrations for defined periods can be computed.

A1.4.6 Buildings and obstacles

The localised effects of building or obstacles (e.g. noise barriers, forests) on dispersion can be modelled in GRAL. Such features can be taken into account in GRAL simulations either using a simple diagnostic model or an advanced prognostic microscale wind field model that is automatically launched whenever buildings are added to the model domain. In general, the recommendation is to use the prognostic wind field model. In cases of very large model domains, for which the prognostic option would lead to very long computation times, the diagnostic model is an alternative. Some specific techniques are being developed and tested to model the effects of obstacles such as noise barriers and sunken tunnel portals.

A1.4.7 Vehicle-induced turbulence

GRAL has no specific algorithms to simulate vehicle-induced turbulence. The effect is considered through the height of the source which defines the initial starting point of particles. For line sources all particles start within a box with a height of 3 metres, and with a user-defined (road) width.

A1.4.8 Post-processing

The steady-state concentration field for each classified meteorological situation, each source group, and each 'horizontal slice' (output height) is stored as a separate file. A post-processing routine in GRAL combines these individual concentration fields to produce an aggregated concentration field for each required statistic (e.g. annual mean, hourly or daily percentile, hourly or daily maximum, user-defined seasonal time slices and user defined day time slices).

A1.4.9 Outputs

Model predictions are available for each selected height above ground level. The main outputs are:

- Single GRAMM wind fields. Maps and cross-section maps of several wind parameters such as wind speed, wind direction, vertical inclination of wind direction.
- GRAMM wind field statistics for the calculated time period. These include the mean, minimum and maximum wind speed values, and the number of hours above or below a certain wind speed.
- Single GRAL concentration situations: maps of concentration, the desired height above ground and layer thickness must be defined before GRAL starts.
- GRAL concentration statistics aggregating the stored single flow situations. Besides the usual display of mean, max and percentiles there are time slices available, source groups with different emission time histograms, and more.

- Grid operations with special functions (e.g. NO_x conversion) and free definable formulas to combine spatial information of several maps are available for all concentration maps, e.g. to derive daily and hourly statistics from annual means (using empirical derivations). There are no operations available for wind flow maps.

A1.5 Typical model inputs and settings

Table A-1 summarises the main model settings and inputs, the physical effects that these simulate, and the qualitative effects of changes to the most important settings/inputs on model predictions. Appropriate values (or ranges) for the model settings and inputs are also highlighted, with a focus on modelling in urban areas.

Table A-1: GRAMM settings

Parameter	Comments	Range/options	Typical value(s)
GRAMM grid			
Terrain	The nature of the terrain around an emission source can significantly affect meteorology and the dispersion of pollutants.	For complex terrain GRAMM should be used. The GRAMM domain needs to be much larger than the GRAL domain. Typical terrain resolution for a dispersion modelling project may range between 30 metres and 90 metres. When performing a micro-scale simulation, or where terrain is particularly complex, the resolution of the terrain used in GRAMM may be higher (2-10 metres).	Terrain resolution is dependent on the size of the GRAL domain and the complexity of the terrain itself. Values of 30-90 metres may be used for most impact assessments. However, much finer terrain may be used where the resolution of the GRAL domain is also high.
Land use	Physical surface parameters (e.g. roughness length) are used resolve surface-atmosphere transfer of momentum, heat, and mass. For example, the near surface temperature depends on land cover (e.g., sea, forest, barren ground) and soil characteristics.	The user can enter characteristic area-average values or more detailed land used data. European CORINE vector-based shape files can be imported.	For standard applications, area-average values for land surface are frequently used. If detailed land use data are available, a land-use pre-processor derives the roughness length, the albedo, the soil thermal conductivity and diffusivity and soil moisture parameters.
Horizontal grid resolution (m)	Defines the horizontal grid size of the wind field. The horizontal resolution of the wind field predictions in GRAMM is lower than that required for GRAL.	The resolution in GRAMM typically ranges from 50 m where there are large and sudden variations in topography to 300 m where topography is less variable and land use can be sensibly resolved (or where the model domain is larger than 5 km x 5 km. If buildings are included in GRAL, then a higher grid resolution may also be required for GRAMM. However, a microscale resolution of terrain and land use structures in GRAMM is neither necessary nor sensible. Prominent topographic details like causeways and similar embankments are generally better represented by a building object within the GRAL model ² .	Typical ranges are: 200-500 m 50-200 m 25-50 m
Vertical thickness of first layer (m)	Defines the cell height of the lowest layer of the flow field.	In terms of vertical resolution, the lower limit for GRAMM is typically 10 m for the ground layer, spreading in the following layers with a factor 1.20 to 1.35 depending on topography (cross-ref to text on stretching factors).	In most cases 10 m can be used.
Number of vertical layers	Defines the number of vertical meteorological layers in the GRAMM model domain.	The user may enter the number of vertical layers starting from one. The recommended minimum number of vertical layers in GRAMM is 15. This number should be chosen whilst also considering the vertical thickness of the first layer and the vertical stretching factor.	15 or more.

² Braunstein + Berndt GmbH. SoundPLAN – Air Pollution Modules. <http://www.soundplan.eu/english/soundplan-air-pollution/>

Parameter	Comments	Range/options	Typical value(s)
Vertical stretching factor	Defines how quickly cell heights increase with height above the ground. For example, a factor of 1.1 means a cell is 10% higher than the one below it.	The stretching factor has to be chosen in a way, such that the top of the model domain is at least three times as high as the highest elevation in the domain.	Values of 1 and above may be used, but it depends on the range of heights in the project terrain. A smaller value (e.g. 1.05 – 1.4) may be used where the terrain is predominantly flat.
Relative layer height (m)	Defines the top of the model domain relative to the lowest level in the domain.	This value is calculated automatically by the GUI based on the vertical layer values input by the user.	A predominantly flat terrain with, for example, a 10 metre vertical thickness of the first layer, 15 vertical layers and a vertical stretching factor of 1.2 would result in a relative layer height of 730 metres.
GRAMM input			
Reference meteorology	The reference meteorology is used to drive the GRAMM wind field calculations.	The reference meteorology may be data from one meteorological station or a synthetic data file which includes a range of meteorological parameters.	The meteorological data has a one-hour time base.
Wind speed classes	GRAMM and GRAL calculations should use classified wind statistics.	Wind speed classes should be defined in accordance with the spread of the input meteorological data.	Dependant on the input data however typical ranges might be: 0 m/s – 0.5 m/s 0.5 m/s – 1 m/s 1 m/s – 2 m/s....etc The higher wind speeds may be grouped (e.g. 4 m/s – 6 m/s).
Max time step (s)	Defines the amount of time taken to ensure that calculations are done efficiently but stably.	In order to keep the simulations numerically stable the maximum time step needs to be below a critical value. There is no way to compute this critical value beforehand; it is up to the experience of a user to set the upper limit. Complexity of the topography and horizontal grid resolution are the most critical parameters influencing numerical stability of simulations.	10 s for a 200-500 m grid 5 s for a 50-200 m grid 2 s for a 25-50 m grid
Modelling time (s)	Defines the integration time for each dispersion situation.	This time is not necessarily the same as the dispersion time for the GRAL.	It is recommended to use the default value of 3,600 seconds, as GRAMM has been tested thoroughly for this integration time.
Relaxation velocity	These factors are chosen to ensure the numerical stability of the GRAMM simulations.	In general, the smaller the relaxation factors are, the better the numerical stability of the simulations.	0.2 for a 200-500 m grid 0.1 for a 50-200 m grid 0.05 for a 25-50 m grid
Relaxation scalars	As above	As above	As above

Table A-2: GRAL settings

Parameter	Comments	Range/options	Typical value
General			
Dispersion time (s)	Defines the shortest averaging time for concentrations.	Lower threshold of 300 seconds, due to validity of turbulence parameterisation in GRAL.	3,600 seconds for hourly values.
Particles per second	Defines the total number of particles released in each dispersion situation (=particles per second x dispersion time). The higher the number of particles, the smoother the concentrations fields (statistical errors are reduced with increasing number of particles). Computation time increases linearly with the number of particles.	User-defined, and dependant on the size of the model domain and the number and types of sources.	Typical values are 25 (for areas < 250 x 250 m ²) and 1000 (for areas > 20 x 20 km ² and numerous sources). In most cases (few sources, areas < 1 km ²) 100 particles per second are a good choice. In the case of many and different types of sources in the same run (i.e. line and point), you may need to increase the number of particles but keeping in mind run time.
Surface roughness (m)	Defines the roughness length in the whole model domain, unless flow fields from GRAMM are taken as meteorological input, and GRAMM is supplied with corresponding land-use data. The roughness lengths are then taken from the land-use file, which might be inhomogeneous. The roughness length alters the shape of the velocity profile near the surface.	User-defined, but based on typical roughness length values found in the literature.	Typical values are: Open sea – 0.00002 Open flat terrain, few isolated obstacles – 0.03 Low crops, occasional obstacles – 0.10 High crops, scattered obstacles – 0.25 Parkland, bushes, numerous obstacles – 0.50 Regular large obstacle coverages (suburb, forest) – 0.5 to 1.0.
Buildings	Buildings and obstacles (e.g. noise barriers, forests) can be taken into account in GRAL simulations either using a simple diagnostic or an advanced prognostic microscale wind field model.	When including buildings in a model run it is important to consider the implications on run time due to the finer spatial resolution that is required.	In general, it is recommended to use the prognostic wind field model. In cases of very large model domains for which the prognostic option would lead to very long computation times, the diagnostic model is an alternative.
Concentration grids			
Horizontal grid resolution (m)	Defines the horizontal grid size of the concentration grid.	User defined based on the size of the model domain and number and type of sources.	The lower bound for the horizontal grid size is 2 m, and there is no upper bound. When buildings are not included in the model the typical horizontal resolution in the model is 10-20 m. When buildings are included in the model a higher horizontal resolution is required, and this is typically 2-5 m.

Parameter	Comments	Range/options	Typical value
Vertical dimension of concentration layers (m)	Defines the vertical extension of the concentration grid.	The vertical resolution in GRAL is typically 0.5-1.0 m for the ground layer in the presence of buildings with a stretching factor of 1.05 (cross-ref to text on stretching factors). Without buildings, a resolution of 2 m for the ground layer and a stretching factor of 1.1 for the following layers is often used. The counting grid can be chosen in a different resolution.	In most cases 10 m can be used for the ground layer.
Number of horizontal slices	Defines the number of horizontal GRAL grids or concentration grids.	A maximum of 10 horizontal grids can be defined.	For typical air quality assessments, 1 will be most common however, if multiple layers at different heights (e.g. elevated receptors) is required, this is an option.
Heights above ground (m)	Defines the height above ground for each concentration grid. The height is defined at the centre of the vertical extension.	In specific reference to the GRAL model, a height of 3 m represents concentrations effectively at 'ground level'. In the GRAL model, 0 m is the direct boundary layer which contains boundary conditions not appropriate for accurate concentration predictions.	In most cases 3 m can be used to represent ground-level concentrations.
Internal flow field grid			
<i>Whenever the diagnostic or prognostic option is selected to take buildings into account, additional input fields define the parameters for the microscale flow field model of GRAL.</i>			
Horizontal grid resolution (m)	Defines the horizontal grid size for the microscale flow field.	This field is automatically updated by the GUI to reflect the horizontal grid resolution chosen in the 'Concentration Grid' window (see notes in section above this one). In some cases, it might be advisable to use different grid sizes for the flow and concentration grids. For example, if high stack emissions are still influenced by large buildings, the flow field grid needs to be small, but the concentration grid should take much larger values to avoid statistical errors.	See 'Horizontal grid resolution' in the 'Concentration grids' section above.
Vertical thickness of first layer (m)	Defines the cell height of the lowest layer of the flow field.	User defined dependant on the size of the model domain.	Depending on the size of the model domain, typical values are 1-2 m.
Vertical stretching factor	Defines how quickly cell heights increase with height above ground. A factor of 1.1 means a cell is 10 % higher than the previous lower one.	It should be noted that rather large cell heights may occur in the presence of mountainous terrain, therefore it is generally recommended to use stretching factors close to 1.0 as long if computationally practical.	1 or close to 1.

Parameter	Comments	Range/options	Typical value
Number of cells in z direction	Defines the number of cells used in the flow field model. The total number of cells in the z-direction is automatically defined within the model (the user cannot change it). The model height should be high enough, such that about 90% of the cross area at any vertical slice through the model domain is without obstacles.	Automatically defined by the model.	Normally defined as 40 cells by the model.
Relative layer height (m)	The number of cells in z-direction together with the vertical thickness of first layer and the vertical stretching factor define the relative model height	Automatically defined by the model.	Determined by the model based on the number of cells in the z-direction, the vertical thickness of the first layer and the vertical stretching factor.
Minimum iterations	Defines the minimum number of iterations performed.	Increasing this number may significantly increase computational times.	-
Maximum iterations	Defines the maximum number of iterations that shall be performed.	Decreasing this number may significantly decrease computational times. However, simulated flow fields may then be far from meeting steady-state conditions.	-
Roughness of building walls	Defines the roughness length of the walls of all obstacles within the model domain.	-	-

Appendix B

Model evaluation principles

B1 Model uncertainty

Atmospheric dispersion models simplify the various complex processes involved in determining ground-level concentrations (GLCs) of pollutants, and are powerful tools for the assessment and forecasting of pollutants concentration in the atmosphere.

As models are increasingly used for policy support their evaluation is becoming an important issue. Where the results from a dispersion model are to be used to support a regulatory decision (and/or expenditure on pollution-control measures), it is essential to provide a measure of the model uncertainty (D'Abreton, 2009). This information about uncertainties associated with modelling results will be as important as modelling results.

Model uncertainty is composed of model chemistry/physics uncertainties, data uncertainties, and stochastic uncertainties. In addition, there is inherent uncertainty in the behaviour of the atmosphere, especially on shorter time scales due to the effects of random turbulence (D'Abreton, 2009). Defra (2016) states that the predicted results from a dispersion model may differ from measured concentrations for a large number of reasons, including:

- Estimates of background concentrations.
- Meteorological data uncertainties.
- Uncertainties in source activity data such as traffic flows and emission factors.
- Model input parameters such as roughness length.
- Uncertainties associated with monitoring data.

The main sources of uncertainty in dispersion models (with respect to traffic on surface roads) and their effects are summarised in Table B-1. In reality, the differences between modelled and monitored results are likely to be a combination of all of these aspects.

Model set-up and input data should be reviewed carefully to reduce uncertainties in predictions. Common improvements that can be made to a model are:

- Checks on traffic data.
- Checks on road widths.
- Check of distance between sources and monitoring sites, as represented in the model.
- Consideration of speed estimates on roads, in particular at junctions.
- Consideration of source type, such as open roads and street canyons.
- Checks on the monitoring data.

Model performance is often assessed in terms of annual average concentrations. Whilst this is important, it is possible that good model performance can result by chance, and this approach does not test whether the model is predicting the correct concentrations for the right reasons. To some extent this was observed in the evaluation of model performance for the M4 East and New M5 projects (Boulter et al., 2015; Manansala et al., 2015). This is an issue that has been acknowledged elsewhere, and where continuous monitoring data are available there is an opportunity to assess model performance in more detail (Bull, 2011).

Table B-1: Summary of main sources of modelling uncertainty (adapted from D'Abreton, 2009)

Source	Effects
Oversimplification of physics in model	Various effects can lead to model under-prediction and over-prediction. Errors are greater in Gaussian plume models, which do not include the effects of non-steady-state meteorology (<i>i.e.</i> spatially- and temporally-varying meteorology).
Errors in emissions data	Predicted concentrations are proportional to emission rates, and in the case of road transport there are many potential sources of error/uncertainty in emission models. These include, for example, the selection of a vehicle sample in the model that is unrepresentative of the local fleet, errors during testing, errors during model formulation, incorrect input data, and user error.
Meteorological data uncertainties	Wind direction affects direction of plume travel. Wind speed affects plume rise and dilution of plume, resulting in potential errors in distance of plume impact from source, and magnitude of impact.
Errors in stability estimates	Gaussian plume models use estimates of stability class, and 3-D models use explicit vertical profiles of temperature and wind (which are used directly or indirectly to estimate stability class for Gaussian models). In either case, errors in these parameters can cause either under prediction or over prediction of ground-level concentrations.
Errors in dispersion coefficients	Most Gaussian models use modifications of the dispersion coefficients derived experimentally by Pasquill in a rural area of fairly level, open terrain and for relatively moderate plume travel distances. Pasquill coefficients could be in error by plus or minus 25 percent, especially when used for non-level, complex terrain and for large distances ranging up to 50 kilometres or more.
Errors in mixing height estimates	The mixing height determines the volume of air any emitted pollutant can be mixed through by turbulence. Errors in the mixing height may therefore result in errors in short term concentration predictions.
Errors in outlet temperature	Temperature affects plume buoyancy.
Estimates of background concentrations	There is often a significant background (non-modelled) contribution to the concentration of a pollutant at a given location, but typically the background is characterised quite poorly. For example, there can be an insufficient number of monitoring sites to enable the temporal and spatial variation of background pollution to be understood.
Inherent uncertainty	Models predict 'ensemble mean' concentrations for any specific set of input data (say on a 1-hour basis), <i>i.e.</i> they predict the mean concentrations that would result from a large set of observations under the specific conditions being modelled. However, for any specific hour with those exact mean hourly conditions, the predicted ground-level concentrations will not exactly match the actual pattern of ground-level concentrations, due to the effects of random turbulent motions and random fluctuations in other factors such as temperature.

Models are more reliable for estimating longer time-averaged concentrations than for estimating short-term concentrations at specific locations and models are reasonably reliable in estimating the magnitude of highest concentrations occurring sometime, somewhere within an area.

B2 Measurement uncertainty

Ideally, the comparison between model predictions and observed concentrations needs to take into account the uncertainty inherent in both sets of data. However, a detailed treatment of measurement uncertainty was outside the scope of this study.

B3 Evaluation methods and metrics

Model evaluation is in general a complex procedure involving different steps (scientific evaluation, code verification, model validation, sensitivity analysis etc.). Models applied for regulatory air quality assessment are commonly evaluated on the basis of comparisons against observations. This element of the model evaluation process is also known as operational model evaluation or statistical performance analysis, since statistical indicators and graphical analysis are used to determine the capability of an air quality model to reproduce measured concentrations. Although the comparison between modelled and observed concentrations cannot give a thorough insight into the properties of the model, it is seen as a good first step in the evaluation of model performance (Thunis et al., 2012).

Model evaluation should comprise five components (D'Abreton, 2009):

- Peer review - evaluates whether the assumptions, methods, and conclusions derived from environmental models are based on sound scientific principles.
- Quality assurance - data quality assessments are a key component of the QA plan for models.
- Sensitivity analysis - a model's sensitivity describes the degree to which the model result is affected by changes in a selected input parameter.
- Uncertainty analysis - is the term to describe incomplete knowledge about specific factors, parameters or models.
- Model performance using statistical tests (see below).

Model validation results can be presented in tabular form, showing results of statistical tests, or in graphical form, showing comparisons between modelled and measured concentration. Various statistical tests and metrics have been developed to assess model performance, and some of these are summarised in Table B2. These have been proposed for different fields of application (e.g. meteorology, air quality, hydrology), different goals (forecasting, study of specific episodes), or different types of application. There is, however, no single statistic that encapsulates all aspects of interest. It is generally recommended that multiple performance indicators are used, regardless of the model application, since each one has advantages and disadvantages. However, although statistical metrics provide insight into model performance in general, they give little or no information on model weakness and cannot identify whether the modelled concentrations are correct for the right or wrong reason. They do not confirm whether model results have reached a sufficient level of quality for a given application (e.g. policy support) (Thunis et al., 2012).

Several model evaluation metrics can be calculated quickly using the modStats function in Openair, and these are also shown in Table B-2.

Table B-2: Model performance metrics and criteria

Metric	Formula	Comments	Ideal value	Criteria for acceptable model performance		Included in Openair
				Wind speed	Air pollutants	
Fraction of predictions within a factor of two (FAC2)	$0.5 \leq \frac{P_i}{O_i} \leq 2.0$	The fraction of modelled values within a factor of two of the observed values.	1.0	No criterion	>0.5 (Derwent et al., 2010)	Yes
Ratio of means (ROM)	$ROM = \frac{1}{N} \sum_{i=1}^N \frac{\bar{P}}{\bar{O}}$	Ratio of the predicted and observed values. A ratio of greater than 1 indicates an average over-prediction and a ratio of less than 1 indicates an average under-prediction.	1.0	No criterion	No criterion	No
Correlation coefficient (r)	$r = \frac{\sum_{i=1}^N (O_i - \bar{O})(P_i - \bar{P})}{O_{std} \times P_{std}}$	Measures the strength of the linear relationship between predicted and observed data. If there is perfect linear relationship with positive slope between the two variables, $r = 1$. If there is a perfect linear relationship with negative slope between the two variables $r = -1$. A correlation coefficient of 0 means that there is no linear relationship between the variables. This statistic can be particularly useful when comparing a large number of data points. Willmott (1982) discourages the use of r because it does not consistently relate to the accuracy of predictions.	1.0	No criterion, but significance level could be used	No criterion, but significance level could be used	Yes
Mean bias (MB)	$MB = \frac{1}{N} \sum_{i=1}^N (P_i - O_i)$	Averages the difference (predicted - observed) over all pairs in which the observed values are greater than zero. This is used to make statements about the absolute bias in the model simulation. Positive values indicate that the model prediction exceeds the observation, whereas negative values indicate an under-prediction by the model. A mean bias of zero indicates that the model over-predictions and under-predictions cancel each other out. The mean bias provides a good indication of the mean over or under estimate of predictions. Mean bias in the same units as the quantities being considered.	Zero	< ±0.5 m/s (Teschke et al., 2002)	No criterion	Yes

Metric	Formula	Comments	Ideal value	Criteria for acceptable model performance		Included in Openair
				Wind speed	Air pollutants	
Geometric Mean Bias (GMB)	$GMB = \exp\left(\frac{1}{N} \sum_{i=1}^N \ln O_i - \frac{1}{N} \sum_{i=1}^N \ln P_i\right)$	This is a measure of mean bias, and indicates only systematic errors.	1.0	No criterion	0.7 < GMB < 1.3 (Chang and Hanna, 2005)	No
Normalised mean bias (NMB)	$NMB = \frac{\sum_{i=1}^N (P_i - O_i)}{\sum_{i=1}^N (O_i)} \times 100\%$	This statistic averages the difference (predicted - observed) over the sum of observed values. It is a useful model performance indicator because it avoids over-inflating the observed range of values. The normalised mean bias is useful for comparing pollutants that cover different concentration scales and the mean bias is normalised by dividing by the observed concentration.	Zero	No criterion	±20% (Derwent et al., 2010) 50% for PM ₁₀ (Thunis et al., 2012)	Yes
Root mean square error (RMSE)	$RMSE = \left[\frac{1}{N} \sum_{i=1}^N (O_i - P_i)^2 \right]^{1/2}$	RMSE is used to define the average error or uncertainty of a model. The units of RMSE are the same as those of the quantity compared. The smaller the RMSE the greater model the precision. RMSE is a good overall measure of model performance. However, large errors are weighted heavily (due to squaring).	Zero	<2 m/s (Teschke et al., 2002)	No criterion	Yes
RMSE normalised by standard deviation of observations (RMSE/σ)	$\frac{RMSE}{O_{std}} = \frac{\left[\sum_{i=1}^N (O_i - P_i)^2 \right]^{1/2}}{\left[\sum_{i=1}^N (O_i - \bar{O})^2 \right]^{1/2}}$	The RMSE normalised by the standard deviation of the observations. A value of less than or equal to one for this indicator means that the model is a better predictor of the observations compared with the mean of the observations.	Zero	No criterion	No criterion	No

Metric	Formula	Comments	Ideal value	Criteria for acceptable model performance		Included in Openair
				Wind speed	Air pollutants	
Fractional bias (FB)	$FB = 2 \frac{\overline{O} - \overline{P}}{\overline{O} + \overline{P}}$	The fractional bias (FB) is used to identify if a model shows a systematic tendency to over- or under-predict. Negative values suggest a model over-prediction and positive values suggest a model under-prediction. FB varies between -2 and +2. FB values of ± 0.67 correspond to a prediction within a factor of 2. Fractional bias is a useful indicator because equally weights positive and negative bias. The main disadvantage is that the predicted concentration is stated in both the numerator and denominator.	Zero	Within ± 0.67 <0.3 (VDI (2005) and Hanna et al. (2004))	Within ± 0.67	No
Mean error (ME) or mean gross error (MGE)	$ME = \frac{1}{N} \sum_{i=1}^N P_i - O_i $	Averages the absolute value of the difference (predicted - observed) over all pairs in which the observed values are greater than zero. It is similar to mean bias except that the absolute value of the difference is used so that the error is always positive. The mean gross error provides a good indication of the mean error regardless of whether it is an over or under estimate. Mean gross error is in the same units as the quantities being considered.	Zero	No criterion	No criterion	Yes
Geometric Variance (GV)	$GV = \exp \left(\frac{1}{N} \sum_{i=1}^N (\ln O_i - \ln P_i)^2 \right)$	Measures scatter and reflects both systematic and unsystematic (random) errors,	1.0	No criterion	1.6 (equivalent to factor-of-two bias) (Chang and Hanna, 2005)	No
Normalised mean error (NME) or normalised mean gross error (NMGE)	$NME = \frac{\sum_{i=1}^N P_i - O_i }{\sum_{i=1}^N (O_i)} \times 100\%$	Used as a normalisation of the mean error to facilitate a range of concentrations. This statistic averages the difference (predicted - observed) over the sum of observed values. Normalised mean error is a useful because it avoids over-inflating the observed range of values. The normalised mean gross error further ignores whether a prediction is an over- or under-estimate.	Zero	No criterion	0.5 (equivalent to factor-of-two bias) (Chang and Hanna, 2005)	Yes

Metric	Formula	Comments	Ideal value	Criteria for acceptable model performance		Included in Openair
				Wind speed	Air pollutants	
Skill_r	$Skill_r = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2}}{O_{std}}$	This is the ratio of RMSE to observed standard deviation. A model is considered to be predicting with skill if the RMSE is less than the standard deviation of the observations (Skill_r <1).	Zero	<1.0	<1.0	No
Skill_v (skill variance)	$Skill_v = \frac{P_{std}}{O_{std}}$	A model is considered to be predicting with skill if the standard deviations of the predictions and observations are the same (Skill_v = 1).	1.0	No criterion	No criterion	No
Coefficient of efficiency (COE)	$COE = 1.0 - \frac{\sum_{i=1}^N P_i - O_i }{\sum_{i=1}^N O_i - \bar{O} }$	A perfect model has a COE of 1.0. A value of zero implies that the model is no more able to predict the observed values than does the observed mean. Therefore, since the model can explain no more of the variation in the observed values than can the observed mean, such a model can have no predictive advantage. For negative values of COE, the model is less effective than the observed mean in predicting the variation in the observations.	1.0	No criterion	No criterion	Yes

Metric	Formula	Comments	Ideal value	Criteria for acceptable model performance		Included in Openair
				Wind speed	Air pollutants	
Index of agreement (IOA)	$IOA = \begin{cases} 1.0 - \frac{\sum_{i=1}^N P_i - O_i }{\sum_{i=1}^N O_i - \bar{O} }, \text{ when } \sum_{i=1}^N P_i - O_i \leq c \sum_{i=1}^N O_i - \bar{O} \\ c \frac{\sum_{i=1}^N P_i - \bar{O} }{\sum_{i=1}^N P_i - O_i } - 1.0, \text{ when } \sum_{i=1}^N P_i - O_i > c \sum_{i=1}^N O_i - \bar{O} \end{cases}$	Values range between -1 and +1 with values approaching +1 representing better model performance. An IOA of 0.5 indicates that the sum of the error magnitudes is one half of the sum of the observed-deviation magnitudes. A value of zero signifies that the sum of the magnitudes of the errors and the sum of the observed-deviation magnitudes are equivalent. When IOA = -0.5, it indicates that the sum of the error-magnitudes is twice the sum of the perfect model-deviation and observed-deviation magnitudes. Values of IOA near -1.0 can mean that the model estimated deviations are poor estimates of the observed deviations; but, they also can mean that there simply is little observed variability — so some caution is needed when the IOA approaches -1. (c = 2).	1.0	≥0.6 (USEPA, 2009)	No criterion	Yes

Definitions:

- i* = the number of observation compared, 1,2, 3 N
N = total number of observations compared (or pairs of data)
O_i = observed (measured) value for the i-th hour
P_i = predicted (modelled) value for the i-th hour
 \bar{O} = average of all observed values
 \bar{P} = average of all predicted values
O_{std} = standard deviation of observed values
P_{std} = standard deviation of predicted values
C_H = Hth highest concentration
 \bar{C} = mean of the top *H*-1 concentrations

Sources: Hanna, 1989; Hurley, 2000; Hurley et al., 2002; USEPA, 2007; Thunis et al., 2012; Defra, 2016; Carslaw (2015); Willmott (1982); Chang and Hanna (2005)

B4 Guidance

The guidance on model evaluation in Australia is limited. Some resources on guidance from overseas are summarised briefly below, and in some cases these have been consulted in this study.

- United Kingdom
 - *Evaluating the Performance of Air Quality Models* (Derwent et al., 2010).
 - Defra Technical Guidance - LAQM.TG(16) - for air quality assessments, which includes sections on model verification (Defra, 2016).
 - *DMRB Air Quality Model Verification – Good Practice Guide* (Bull, 2011).
- European Union
 - A detailed review of guidance on the use of models, and model evaluation, for the EU Air Quality Directive 2008/50/EC was compiled by Denby et al. (2010). The work was conducted within the framework of the FAIRMODE (Forum for Air Quality Modeling in Europe) project³.
 - COST Action 728 on standardised model evaluation protocol for meso-scale meteorological models.
 - COST Action 732 on quality assurance of micro-scale (obstacles resolving) meteorological models.
 - The Air4EU project included validation and uncertainty analysis for models covering a broad spatial scale.
- United States
 - *Guidance on the Development, Evaluation, and Application of Environmental Models* (USEPA, 2009).
 - ASTM D6589-05: *Standard Guide for Statistical Evaluation of Atmospheric Dispersion Model Performance* (ASTM, 2015).

³ <http://fairmode.ew.eea.europa.eu/>

Appendix C

Review of GRAL validation studies

C1 Model validation studies

C1.1 Studies cited in GRAL documentation

The GRAMM-GRAL system has been validated in numerous studies, as described in the model documentation⁴. These studies have used datasets for:

- Multiple countries (USA, Norway, Denmark, Germany, Sweden, Austria, Japan, Finland).
- Multiple source types (power plant stacks, elevated tracers, ground-level tracers, urban roads, street canyons, parking lots, tunnel portals).
- Different terrain types (flat and complex).
- Varying meteorological conditions (high/low wind speeds, stable/unstable atmospheric conditions etc.).

The studies are summarised in Table C-1.

In several cases the performance of GRAMM-GRAL has been shown to be at least as good as that of other models, including ADMS, AERMOD, CALINE, ISC and AUSTAL. However, it is worth noting that all models are constantly being improved, and the performance of each model will change with time.

In the case of GRAL, the documentation on validation is periodically updated. For each update to the model, the validation tests are repeated using 29 different data sets (including those in Table C-1).

The GRAL system had not been used extensively in Australia prior to WestConnex, and there had been no Australia-specific validation or verification work.

For the road studies, the performance of GRAL compared with that of other models is summarised in Table C-2.

⁴ http://app.luis.steiermark.at/berichte/Download/Fachberichte/LU_08_14_GRAL_Documentation.pdf

Table C-1: Validation studies identified in GRAL documentation

Country	Dataset	Type of source	Type of study	Relevant to study
Studies cited for compliance with the Austrian Guideline RVS 04.02.12				
USA	California (CALTRANS 99 study)	Line source (road) ^(a)	Tracer gas	✓
Austria	Biedermannsdorf (A2 Highway)	Line source (road) ^(b)	Emission modelling and ambient measurement	✓
Austria	Ehrentalerberg	Tunnel portal	Tracer gas	×
Austria	Kaisermuehlen	Tunnel portal	Tracer gas	×
Additional validation studies				
USA	Indianapolis (Perry K)	Point (power plant stack)	Tracer gas	×
USA	Illinois (Kincaid plant)	Point (power plant stack)	Tracer gas	×
Norway	Lillestroem	Simulated point (high mast)	Tracer gas	×
USA	Nebraska (prairie grass)	Simulated point (low mast)	Tracer gas	×
Denmark	Copenhagen	Simulated point (high mast)	Tracer gas	×
USA	Idaho	Simulated point (low mast)	Tracer gas	×
Austria	Raaba	Simulated point (low mast)	Tracer gas	×
Austria	Gratkorn	Point (paper mill stack)	Stack emission and ambient measurement	×
USA	Idaho Falls	Line source, with and without noise barrier	Tracer gas	✓
Finland	Elimaeki	Line source (road)	Emission modelling and ambient measurement	✓
Germany	Hannover, Goettinger Strasse	Line source (road)	Emission modelling and ambient measurement	✓
Germany	Berlin (Frankfurter Allee)	Line source (road)	Emission modelling and ambient measurement	✓
Sweden	Stockholm (Hornsgatan)	Line source (road, street canyon)	Emission modelling and ambient measurement	✓
-	U-shaped building	-	Wind tunnel	×
Austria	Vienna (parking lot)	Area source (tracer gas)	Tracer gas	×
Germany	Uttenweile (pig pen)	Area source (tracer gas)	Tracer gas	×
Denmark	Roager (pig pen)	Point sources (tracer gas)	Tracer gas	×
USA	Idaho (Experimental Organically Cooled Reactor study)	Point sources (tracer gas)	Tracer gas	×
USA	Texas/Kansas (American Gas Association experiments)	Point source (gas compressor station stack)	Tracer gas	×
USA	Alaska (North Slope Tracer Study)	Turbine stack	Tracer gas	×
Japan	Ninomiya tunnel	Tunnel portal	Tracer gas	×
Japan	Hitachi tunnel	Tunnel portal	Tracer gas	×
Japan	Enrei tunnel	Tunnel portal	Tracer gas	×

(a) 4-lane highway without accompanying buildings; mostly low wind speeds

(b) 4-lane highway with a noise abatement wall; moderate wind speeds.

Table C-2: Summary of model performance for roads in GRAL documentation

Dataset	Model	NMSE	FB	Mean deviation
GRAL				
California (CALTRANS 99 study)	GRAL	+0.5	0.0	-
Biedermannsdorf (A2 Highway)	GRAL	-	-	-0.2 to +0.2
Idaho Falls	GRAL	+2.5	0	-
Elimaeki	GRAL	+0.1	+0.2	-
Hannover, Goettinger Strasse	GRAL (level 2) ^(a)	+0.7	-0.3	-
Berlin (Frankfurter Allee)	GRAL (level 2)	+1.4	-0.1	-
Stockholm (Hornsgatan)	GRAL	+0.8	+0.1	-
Other models				
California (CALTRANS 99 study)	ADMS Roads	+0.2	+0.1	-
	AERMOD	+0.3	+0.1	-
	CALINE	+0.9	+0.2	-
	RLINE	+0.3	+0.1	-
	MISKAM	+0.5	+0.3	-
Idaho Falls	ADMS Roads	+1.2	+0.4	-
	AERMOD	+1.3	+0.3	-
	CALINE	+2.0	+0.4	-
	RLINE	+1.0	+0.2	-
Elimaeki	CAR FMI	+0.2	-0.1	-
Hannover, Goettinger Strasse	MISKAM	-	0	-
	AUSTAL 2000	-	-0.4	-
Berlin (Frankfurter Allee)	ADMS Roads	+3.1	+1.0	-
	LASAT	+2.2	+0.4	-
	OSPM	-	-0.1	-
	MIMO	-	-0.3	-
Stockholm (Hornsgatan)	ADMS Roads	+1.2	+0.1	-
	LASAT	+0.7	+0.4	-
	OSPM	-	+0.1	-
	MIMO	-	-0.1	-

Key

	Model performance within acceptable range (as stated in GRAL documentation)
	Model performance outside acceptable range (as stated in GRAL documentation)

(a) The 'level 2' approach involved the use of the prognostic met model.

C1.2 Other studies in the literature

The scientific literature contains many studies involving the evaluation of road dispersion models. Although these cannot be covered comprehensively here, a cross-section of publications has been considered to illustrate the typical results for studies that are not presented in the GRAL documentation. However, some of the datasets and models used are the same as in Table C-2. Table C-3 summarises the results from some of the studies in the literature, again with reference to model evaluation metrics. The emphasis is on short-term comparisons (one hour or less) at sites near roads. Some additional

context on these studies is provide below, but clearly not all details of model set-up can be reproduced here.

Levitin et al. (2005) compared two atmospheric dispersion models (CALINE 4 and CAR-FMI) against the results of a measurement campaign near a major road at Elimäki in southern Finland. The concentrations of NO_x, NO₂ and O₃ were measured simultaneously at three locations and at three heights (3.5, 6 and 10 metres) on both sides of the road (at distances from the road of 17 and 34 metres). Table C-3 only shows the results for NO_x at a distance of 17 metres (as it was closer to the road) and a height of 3.5 metres. The agreement of measured and predicted datasets was good for both models. The performance of both models tended to deteriorate as the wind speed decreased, and as the wind direction approached a direction parallel to the road.

Carslaw (2011) summarised the evaluation of models used for the assessment of urban air quality. Although this focussed on hourly mean concentrations, the evaluation of hourly predictions was considered for a subset of models and receptor locations. The aim of the report was to provide information to inform the future use of air quality models in the UK. The author used a number of different datasets and models, and only a sub-set for roadside sites is presented here.

Stocker et al. (2013) summarised the results from a model inter-comparison study undertaken with four road dispersion models and two observational SF₆ tracer gas datasets (Idaho Falls and CALTRANS 99), with concentrations paired in time and space. The Idaho Falls experiment was undertaken in a relatively controlled environment, where the release was from a line source with no traffic. The absence of vehicle-induced turbulence permitted a detailed analysis of lateral dispersion. In the CALTRANS experiment the SF₆ was released from within a real-world traffic stream, giving a more realistic situation. Again, a sub-set of the results is summarised here. With the time-space pairing restriction removed, quantile-quantile plots (not shown) revealed an excellent level of agreement, particularly for the ADMS-Roads, AERMOD (volume source) and RLINE models.

Srimath et al. (2017) evaluated the performance of the CAR-FMI model against measurements at an urban roadside site in London, for 2003-2004 and for 2008 (only the NO_x results for the latter are shown in the table). The concentrations of PM_{2.5}, PM₁₀, NO_x and NO₂ were predicted using the roadside dispersion model CAR-FMI, combined with a national U.K. emission model, a meteorological pre-processor, and measured values at urban background stations. The Indices of Agreement (IA) in all the campaigns ranged from 0.68 to 0.78, 0.87, from 0.70 to 0.80, and from 0.61 to 0.83 for PM_{2.5}, PM₁₀, NO_x and NO₂, respectively.

The small sample of studies identified here have shown how model performance can vary substantially. For example, R² values ranged from 0.22 to 0.77, and the Fac2 values ranged from 0.51 to 0.95. However, the latter are typically in the range of 0.60-0.80.

Table C-3: Summary of model performance for roads (literature)

Location/model	Reference	Bias	NMSE	FB	R / R ²	IOA	Fac2
ADMS Roads							
Idaho Falls ^(a)	Stocker et al. (2013)	-1.74	1.16	-0.37	0.88 / 0.77	-	0.69
CALTRANS 99 ^(b)	Stocker et al. (2013)	-0.13	0.20	-0.09	0.78 / 0.61	-	0.85
ADMS Urban							
Camden, Shaftesbury Avenue	Carslaw (2011)	-	-	-	0.55 / 0.30	-	0.74
Croydon, George Street	Carslaw (2011)	-	-	-	0.75 / 0.56	-	0.83
Ealing, Acton Town Hall	Carslaw (2011)	-	-	-	0.66 / 0.44	-	0.80
Haringey Town Hall	Carslaw (2011)	-	-	-	0.68 / 0.46	-	0.83
Lambeth, Brixton Road	Carslaw (2011)	-	-	-	0.63 / 0.40	-	0.51
Lewisham, New Cross	Carslaw (2011)	-	-	-	0.64 / 0.41	-	0.70
Tower Hamlets, Mile End Road	Carslaw (2011)	-	-	-	0.62 / 0.38	-	0.60
AERMOD (area)							
Idaho Falls ^(a)	Stocker et al. (2013)	-1.58	1.26	-0.33	0.82 / 0.67	-	0.58
CALTRANS 99 ^(b)	Stocker et al. (2013)	-0.18	0.31	-0.13	0.72 / 0.52	-	0.76
AERMOD (volume)							
Idaho Falls ^(a)	Stocker et al. (2013)	-1.75	1.26	-0.37	0.84 / 0.71	-	0.58
CALTRANS 99 ^(b)	Stocker et al. (2013)	-0.21	0.28	-0.15	0.77 / 0.59	-	0.78
CALINE							
Elimäki, Finland ^(c)	Levitin et al. (2005)	-	0.21	0.18	0.74 / 0.55	0.84	0.81
Idaho Falls ^(a)	Stocker et al. (2013)	-1.96	1.97	-0.42	0.76 / 0.58	-	0.58
CALTRANS 99 ^(b)	Stocker et al. (2013)	-0.26	0.86	-0.19	0.47 / 0.22	-	0.68
CAR-FMI							
Elimäki, Finland ^(c)	Levitin et al. (2005)	-	0.19	-0.01	0.81 / 0.66	0.88	0.95
Cromwell Rd, London ^(c)	Srimath et al. (2017)	-	-	0.46	0.60 / 0.36	0.70	0.50
KCL Urban CMAQ							
Camden, Shaftesbury Avenue	Carslaw (2011)	-	-	-	0.57 / 0.32	-	0.78
Croydon, George Street	Carslaw (2011)	-	-	-	0.51 / 0.26	-	0.60

Location/model	Reference	Bias	NMSE	FB	R / R ²	IOA	Fac2
Ealing, Acton Town Hall	Carslaw (2011)	-	-	-	0.53 / 0.28	-	0.70
Haringey Town Hall	Carslaw (2011)	-	-	-	0.60 / 0.36	-	0.75
Lambeth, Brixton Road	Carslaw (2011)	-	-	-	0.61 / 0.37	-	0.75
Lewisham, New Cross	Carslaw (2011)	-	-	-	0.60 / 0.36	-	0.64
Tower Hamlets, Mile End Road	Carslaw (2011)	-	-	-	0.58 / 0.34	-	0.72
RLINE							
Idaho Falls ^(a)	Stocker et al. (2013)	-1.09	0.96	-0.22	0.84 / 0.71	-	0.72
CALTRANS 99 ^(b)	Stocker et al. (2013)	-0.07	0.34	-0.05	0.75 / 0.56	-	0.78

(a) SF₆ tracer, 15-minute sampling

(b) SF₆ tracer, 30-minute sampling

(c) NO_x, 1-hour sampling

Appendix D

Statistical summary of meteorological measurements

D1 Meteorological measurements

Descriptive statistics for the parameters measured at the meteorological stations are provided in Tables D-1 to D-8. These statistics relate to data availability, the distribution of values (e.g. maximum, average, percentiles) and, in the case of wind speed, the percentage of calm winds (average wind speed <0.5 m/s).

The average wind speed and percentage of calm winds are important metrics for summarising the general wind situation at a given location, and for comparing different locations. These values are therefore highlighted in bold.

Table D-1: Summary of meteorological parameters (1-hour average): Concord Oval (2015) (NB: height of 2 metres)

Statistic	Temp. (°C)	Pressure (mmHg)	Rel. humidity (%)	Rainfall (mm)	Solar rad. (W/m ²)	Wind direction (degrees)	Wind speed (m/s)
Availability							
Possible hours	8,760	8,760	8,760	8,760	8,760	8,760	8,760
Valid hours	8,318	8,317	8,230	3,841	8,230	7,854	7,854
Availability (%)	95%	95%	94%	44%	94%	90%	90%
Statistics							
Average value	18.4	1,018.4	69.0	0.0	162.2	-	1.4
Maximum value	42.5	1,038.2	98.6	6.8	1,052.5	-	6.8
2nd highest value	42.0	1,038.2	98.6	4.7	1,034.0	-	6.6
3rd highest value	41.9	1,038.1	98.6	3.5	1,027.0	-	6.5
4th highest value	41.6	1,037.7	98.6	2.0	1,021.6	-	6.4
5th highest value	40.8	1,037.6	98.4	1.2	1,017.0	-	6.4
Minimum value	3.9	996.8	10.2	0.0	0.0	-	0.0
99th percentile	32.4	1,033.9	95.0	0.1	952.7	-	5.0
98th percentile	29.9	1,032.6	94.3	0.0	903.1	-	4.5
95th percentile	27.3	1,030.4	92.7	0.0	763.1	-	3.8
90th percentile	25.0	1,027.5	90.1	0.0	578.0	-	3.2
75th percentile	21.9	1,023.0	82.9	0.0	255.0	-	2.0
50th percentile (median)	18.6	1,018.3	70.5	0.0	1.0	-	1.1
Calm winds (%)	-	-	-	-	-	-	23.1%

Table D-2: Summary of meteorological parameters (1-hour average): St Lukes Park (2015)

Statistic	Temp. (°C)	Pressure (mmHg)	Rel. humidity (%)	Rainfall (mm)	Solar rad. (W/m ²)	Wind direction (degrees)	Wind speed (m/s)
Availability							
Possible hours	8,760	8,760	8,760	8,760	8,760	8,760	8,760
Valid hours	8,720	8,720	8,720	5,051	8,721	8,720	8,720
Availability (%)	100%	100%	100%	58%	100%	100%	100%
Statistics							
Average value	18.0	1,017.6	74.0	0.1	171.7	-	1.3
Maximum value	41.2	1,037.7	100.0	25.2	1,047.6	-	6.4
2nd highest value	41.1	1,037.6	100.0	25.0	1,039.9	-	6.2
3rd highest value	41.0	1,037.6	100.0	19.5	1,033.0	-	6.1
4th highest value	40.7	1,037.3	100.0	17.5	1,027.7	-	6.0
5th highest value	40.1	1,036.9	100.0	15.5	1,025.8	-	6.0
Minimum value	3.5	996.2	11.5	0.0	0.0	-	0.0
99th percentile	31.4	1,033.2	99.9	3.5	952.0	-	4.8
98th percentile	29.1	1,031.9	99.5	1.8	905.9	-	4.5
95th percentile	26.6	1,029.7	98.0	0.5	775.0	-	3.8
90th percentile	24.6	1,026.7	95.6	0.0	592.0	-	3.1
75th percentile	21.6	1,022.2	88.0	0.0	290.0	-	1.9
50th percentile (median)	18.2	1,017.4	75.6	0.0	3.0	-	1.0
Calm winds (%)	-	-	-	-	-	-	27.9%

Table D-3: Summary of meteorological parameters (1-hour average): Sydney Olympic Park (2015)

Statistic	Temp. (°C)	Pressure (mmHg)	Rel. humidity (%)	Rainfall (mm)	Solar rad. (W/m ²)	Wind direction (degrees)	Wind speed (m/s)
Availability							
Possible hours	8,760	-	-	8,760	-	8,760	8,760
Valid hours	8,760	-	-	8,760	-	7,762	7,762
Availability (%)	100%	-	-	100%	-	89%	89%
Statistics							
Average value	17.7	-	-	0.1	-	-	2.6
Maximum value	40.4	-	-	28.2	-	-	8.6
2nd highest value	40.2	-	-	22.8	-	-	8.5
3rd highest value	39.9	-	-	21.8	-	-	8.5
4th highest value	39.9	-	-	17.4	-	-	8.5
5th highest value	39.3	-	-	15.6	-	-	8.4
Minimum value	1.9	-	-	0.0	-	-	0.0
99th percentile	32.9	-	-	3.8	-	-	7.0
98th percentile	30.7	-	-	2.2	-	-	6.4
95th percentile	27.5	-	-	0.4	-	-	5.5
90th percentile	25.3	-	-	0.0	-	-	4.8
75th percentile	21.6	-	-	0.0	-	-	3.7
50th percentile (median)	18.0	-	-	0.0	-	-	2.4
Calm winds (%)	-	-	-	-	-	-	11.1%

Table D-4: Summary of meteorological parameters (1-hour average): Canterbury Racecourse (2015)

Statistic	Temp. (°C)	Pressure (mmHg)	Rel. humidity (%)	Rainfall (mm)	Solar rad. (W/m ²)	Wind direction (degrees)	Wind speed (m/s)
Availability							
Possible hours	8,760	-	8,760	8,760	-	8,760	8,760
Valid hours	8,760	-	8,760	8,760	-	7,807	7,807
Availability (%)	100%	-	100%	100%	-	89%	89%
Statistics							
Average value	17.0	-	71.6	0.1	-	-	3.2
Maximum value	40.2	-	99.7	31.0	-	-	10.1
2nd highest value	40.1	-	99.3	27.0	-	-	10.0
3rd highest value	39.9	-	99.2	26.0	-	-	10.0
4th highest value	39.5	-	99.2	21.8	-	-	9.8
5th highest value	39.4	-	99.1	18.8	-	-	9.6
Minimum value	0.6	-	9.6	0.0	-	-	0.0
99th percentile	31.7	-	99.0	3.0	-	-	8.1
98th percentile	29.2	-	98.3	1.8	-	-	7.5
95th percentile	26.4	-	96.3	0.4	-	-	6.6
90th percentile	24.2	-	94.1	0.0	-	-	5.9
75th percentile	20.9	-	87.6	0.0	-	-	4.6
50th percentile (median)	17.3	-	73.3	0.0	-	-	3.0
Calm winds (%)	-	-	-	-	-	-	7.7%

Table D-5: Summary of meteorological parameters (1-hour average): Rozelle (2015)

Statistic	Temp. (°C)	Pressure (mmHg)	Rel. humidity (%)	Rainfall (mm)	Solar rad. (W/m ²)	Wind direction (degrees)	Wind speed (m/s)
Availability							
Possible hours	8,760	-	8,760	-	8,760	8,760	8,760
Valid hours	8,597	-	8,597	-	8,598	8,443	8,443
Availability (%)	98%	-	98%	-	98%	96%	96%
Statistics							
Average value	17.9	-	72.0	-	165.8	-	1.7
Maximum value	40.7	-	102.0	-	1,108.4	-	16.3
2nd highest value	40.6	-	101.7	-	1,092.7	-	8.2
3rd highest value	40.1	-	101.5	-	1,091.6	-	8.2
4th highest value	40.1	-	101.5	-	1,091.2	-	7.9
5th highest value	39.2	-	101.4	-	1,083.0	-	7.8
Minimum value	4.5	-	15.0	-	-19.7	-	0.0
99th percentile	30.9	-	97.1	-	1,010.4	-	6.0
98th percentile	28.8	-	96.4	-	960.6	-	5.4
95th percentile	26.3	-	95.0	-	813.1	-	4.4
90th percentile	24.2	-	92.5	-	607.7	-	3.7
75th percentile	21.4	-	85.5	-	271.0	-	2.5
50th percentile (median)	18.1	-	73.3	-	-5.3	-	1.2
Calm winds (%)	-	-	-	-	-	-	20.3%

Table D-6: Summary of meteorological parameters (1-hour average): Chullora (2015)

Statistic	Temp. (°C)	Pressure (mmHg)	Rel. humidity (%)	Rainfall (mm)	Solar rad. (W/m ²)	Wind direction (degrees)	Wind speed (m/s)
Availability							
Possible hours	8,760	-	8,760	-	8,760	8,760	8,760
Valid hours	8,672	-	8,672	-	8,666	8,648	8,648
Availability (%)	99%	-	99%	-	99%	99%	99%
Statistics							
Average value	17.7	-	71.2	-	186.6	-	1.7
Maximum value	42.2	-	100.8	-	1,164.7	-	8.2
2nd highest value	41.8	-	100.7	-	1,152.3	-	8.1
3rd highest value	41.7	-	100.6	-	1,151.1	-	8.0
4th highest value	41.3	-	100.6	-	1,147.7	-	8.0
5th highest value	40.7	-	100.5	-	1,143.7	-	7.9
Minimum value	3.0	-	11.2	-	-7.5	-	0.0
99th percentile	33.1	-	99.0	-	1,060.1	-	5.8
98th percentile	30.7	-	98.2	-	991.0	-	4.9
95th percentile	27.4	-	96.6	-	843.1	-	4.0
90th percentile	25.1	-	94.3	-	653.6	-	3.3
75th percentile	21.4	-	86.8	-	317.8	-	2.3
50th percentile (median)	17.9	-	73.4	-	3.9	-	1.5
Calm winds (%)	-	-	-	-	-	-	9.4%

Table D-7: Summary of meteorological parameters (1-hour average): Concord Oval (**November 2016 to February 2017**)

Statistic	Temp. (°C)	Pressure (mmHg)	Rel. humidity (%)	Rainfall (mm)	Solar rad. (W/m ²)	Wind direction (degrees)	Wind speed (m/s)
Availability							
Possible hours	2,880	2,880	2,880	2,880	2,880	2,880	2,880
Valid hours	2,880	2,863	2,868	2,868	2,868	2,860	2,860
Availability (%)	100%	99%	100%	100%	100%	99%	99%
Statistics							
Average value	25.7	1,012.3	65.8	0.1	225.4	-	1.8
Maximum value	46.1	1,025.3	95.8	37.1	1,075.7	-	6.0
2nd highest value	44.9	1,025.2	95.7	23.1	1,050.8	-	5.7
3rd highest value	44.4	1,025.2	95.6	17.0	1,036.0	-	5.6
4th highest value	43.5	1,025.0	95.2	16.3	1,032.0	-	5.5
5th highest value	43.3	1,025.0	95.2	13.7	1,031.8	-	5.5
Minimum value	12.5	992.3	13.8	0.0	0.0	-	0.1
99th percentile	39.5	1,022.8	94.2	2.5	986.2	-	4.8
98th percentile	37.1	1,022.2	92.6	1.0	966.9	-	4.4
95th percentile	34.0	1,021.0	89.4	0.3	913.8	-	4.1
90th percentile	31.6	1,019.6	85.5	0.0	801.7	-	3.6
75th percentile	27.9	1,016.6	78.3	0.0	413.9	-	2.6
50th percentile (median)	25.2	1,012.7	68.1	0.0	32.5	-	1.6
Calm winds (%)	-	-	-	-	-	-	11.3%

Table D-8: Summary of meteorological parameters (1-hour average): St Lukes Park (**November 2016 to February 2017**)

Statistic	Temp. (°C)	Pressure (mmHg)	Rel. humidity (%)	Rainfall (mm)	Solar rad. (W/m ²)	Wind direction (degrees)	Wind speed (m/s)
Availability							
Possible hours	2,880	2,880	2,880	2,880	2,880	2,880	2,880
Valid hours	2,880	2,855	2,855	2,855	2,855	2,837	2,855
Availability (%)	100%	99%	99%	99%	99%	99%	99%
Statistics							
Average value	23.3	1,011.8	70.4	0.1	246.9	-	1.9
Maximum value	43.2	1,024.8	100.0	28.7	1,059.9	-	6.5
2nd highest value	42.8	1,024.7	100.0	20.8	1,057.6	-	6.1
3rd highest value	41.3	1,024.6	100.0	18.8	1,046.4	-	6.0
4th highest value	40.5	1,024.5	100.0	17.3	1,037.1	-	5.9
5th highest value	40.0	1,024.5	100.0	17.3	1,037.0	-	5.9
Minimum value	11.4	991.8	15.7	0.0	0.0	-	0.1
99th percentile	36.6	1,022.4	99.8	2.5	1,002.5	-	5.3
98th percentile	34.5	1,021.7	98.8	1.0	973.8	-	5.0
95th percentile	31.3	1,020.6	95.1	0.0	929.9	-	4.5
90th percentile	28.9	1,019.2	90.8	0.0	815.5	-	3.8
75th percentile	25.2	1,016.1	83.3	0.0	466.2	-	2.7
50th percentile (median)	22.7	1,012.1	73.0	0.0	49.0	-	1.7
Calm winds (%)	-	-	-	-	-	-	14.8%

Appendix E

Results of met model evaluation

E1 Wind speed evaluation

E1.1 Overview

This appendix presents the results from the CALMET and GRAMM performance tests for wind speed, as described in section 4.5.1 (Table 16). In these tests the model predictions were compared with observations (measurements) during the calendar year 2015. All tests were conducted using 1-hour average data, and only the hours with valid data in all three datasets (observations, CALMET and GRAMM) were used. Two series of wind speed tests were conducted:

- Series A. In these tests the reference meteorological data for both CALMET and GRAMM were taken from a single monitoring station (St Lukes Park), and the model predictions were compared with observations at all monitoring stations. The effects of the GRAMM grid resolution and Re-Order function were also tested.
- Series B. In these tests the reference meteorological data were taken from multiple stations, and the model predictions were compared with observations. Firstly, all stations except St Lukes Park were used to provide the reference meteorological data, and then all stations were used. The reference station data were entered directly into CALMET, whereas in GRAMM a synthetic meteorological file with Match-to-Observations was used.

For each test the following results are presented (with some overlap/duplication):

- (a) Time variation plot. The `timeVariation` function in Openair was used to examine model performance in relation to average wind speed by hour of the day, day of the week, and month of the year. Also shown in the plots is the 95% confidence interval on the mean.
- (b) Linear regression plot. Scatter plots and regression analysis are commonly used techniques for model evaluation. The Openair `scatterPlot` function was used for a linear regression analysis. Each plot includes the regression equation, R^2 value, the 1:1 line (solid) and the 1:0.5 and 1:2 lines (dashed). The dashed lines show the predictions that are within a factor of two of the observations (also defined by the FAC2 statistic).
- (c) Quantile-quantile plot. A 'quantile-quantile' plot (Q-Q plot) separately considers the distributions of observations and predictions. Each set of predictions and observations is ranked from highest to lowest. The two ranked sets of data are then plotted as predicted against observed. This plot removes any temporal 'pairing' in the observations and predictions. If the datasets have a similar distribution the plotted values will fall along a 1:1 line. Values above the 1:1 line indicate model overestimation, and values under the 1:1 line indicate model underestimation. A model's general capability to simulate low, average, or high values can therefore easily be visualised.
- (d) Taylor diagram⁵. The Taylor diagram shows how three model performance statistics vary simultaneously. It is explained in more detail below. In this study the `TaylorDiagram` function in Openair was used to produce the plot.
- (e) Statistical metrics. The metrics calculated in Openair using the `modStats` function (see Appendix B) are presented at the bottom of each plot.

Figure E-1 shows an example of a Taylor diagram, which compares model predictions and observations in terms of their centred root-mean-square error (RMSE), the amplitude of their variation (standard

⁵ Details of the diagram can be found at http://www-pcmdi.llnl.gov/about/staff/Taylor/CV/Taylor_diagram_primer.pdf

deviation), and their correlation coefficient (r). The means of the predictions and observations are subtracted from the datasets before the statistics are calculated, so the diagram does not provide information about overall model bias; it only characterises the ‘centred’ error pattern. In this example the performance of five different models (A to E) is compared. The position of each model (letter) in the plot quantifies how closely the model predictions match the observations.

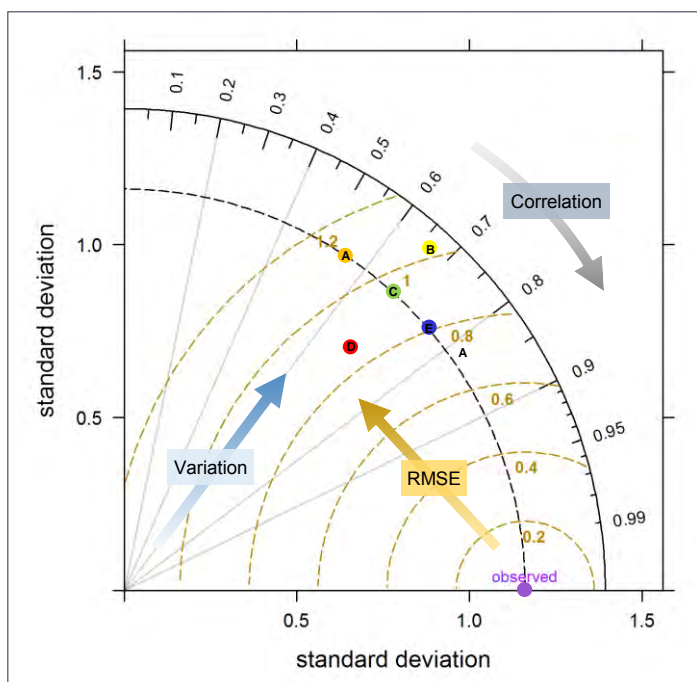


Figure E-1: Example of Taylor diagram

Model predictions that agree well with observations will lie nearest the point marked ‘observed’ on the x-axis. The RMSE for a given model is proportional to the distance to the point on the x-axis identified as ‘observed’. The RMSE for the observations is zero, and the lower the value the better the model performance. The dashed brown contours indicate the RMSE values. The variation (standard deviation) of the model is proportional to the radial distance from the origin in the bottom left-hand corner. The observed variation in this example is around 1.2 units, and to aid interpretation this is shown as a radial dashed black line. Models lying on the dashed black line will have the correct standard deviation (which indicates that the pattern variations are of the right amplitude). The correlation coefficient is shown on the arc, and points that lie closest to the x-axis have the highest correlation. For the observations r is equal to unity. The grey lines represent specific correlation coefficients.

In the Figure, model E has the best performance. Model A has a correlation coefficient of around 0.55, and model E has a correlation coefficient of around 0.75. For models B, C and D the correlation coefficient is around 0.65. Model E has a slightly lower RMSE than model D. The other models have higher RMSE values. Models A, C and E have approximately the same variation as the observations. Model B has too much variation compared with the observations, and model D has too little.

E1.2 Series A (single station reference meteorology)

E1.2.1 CALMET vs GRAMM (tests CT-01 and GM-01)

E1.2.1.1 Between-model comparison

Tests CT-01 and GM-01 involved running CALMET and GRAMM with reference meteorology taken from the St Lukes Park station only and with a grid resolution of 50 metres. The two tests constituted a comparison between CALMET and an initial GRAMM set-up.

Firstly, the predictions were compared with the observations at St Lukes Park, as shown in Figure E-2. Whilst this did not constitute an independent test as such, it illustrated the behaviour of CALMET and GRAMM in relation to their inputs. In this case CALMET exhibited 'perfect' performance, and this is a known characteristic of the model in relation to reference meteorology. Whilst the level of agreement with observations was lower for GRAMM than for CALMET at St Lukes Park (reflecting the meteorological situation modelling approach in GRAMM), the relationship was still very strong ($R^2 = 0.79$).

As far as the performance of GRAMM itself at St Lukes Park is concerned, the following observations have been made:

- The model performance was very similar on all days.
- The model underestimated wind speeds on average. GRAMM typically underestimated the wind speed when the observed wind speed was greater than around 2 m/s.
- GRAMM did not fully represent the diurnal variation in wind speed. The model performance was good between 18:00 and 06:00, and poorest at around 15:00 (when the average underestimation was around 1 m/s).
- The model performance was good in winter (especially April to August), and poorer in summer.

The CALMET and GRAMM predictions for the other evaluation stations are shown in Figures E-3 to E-7. When viewed overall, the performance of CALMET was slightly better than that of GRAMM. However, the performance of both models was poorer at these stations than at St Lukes Park, and for some situations GRAMM performed better than CALMET.

At Sydney Olympic Park (Figure E-3) CALMET showed a good prediction of the average wind speed at night-time, but during the daytime the wind speed was underestimated by almost 2 m/s. The underestimation was also greater during the summer than during the winter. The relationship between the predicted and observed values was moderate ($R^2 = 0.33$), and the regression fit to the data gave an offset on the y-axis of 1.2 m/s. The quantile-quantile plot shows that CALMET overestimated the wind speed at the lowest end of the wind speed distribution (where the value for a quantile in the observations was less than around 1.5 m/s), and underestimated the wind speed for higher quantiles. The GRAMM predictions were systematically lower than those from CALMET (by around 0.8 m/s on average), and showed a poorer temporal agreement with the observations.

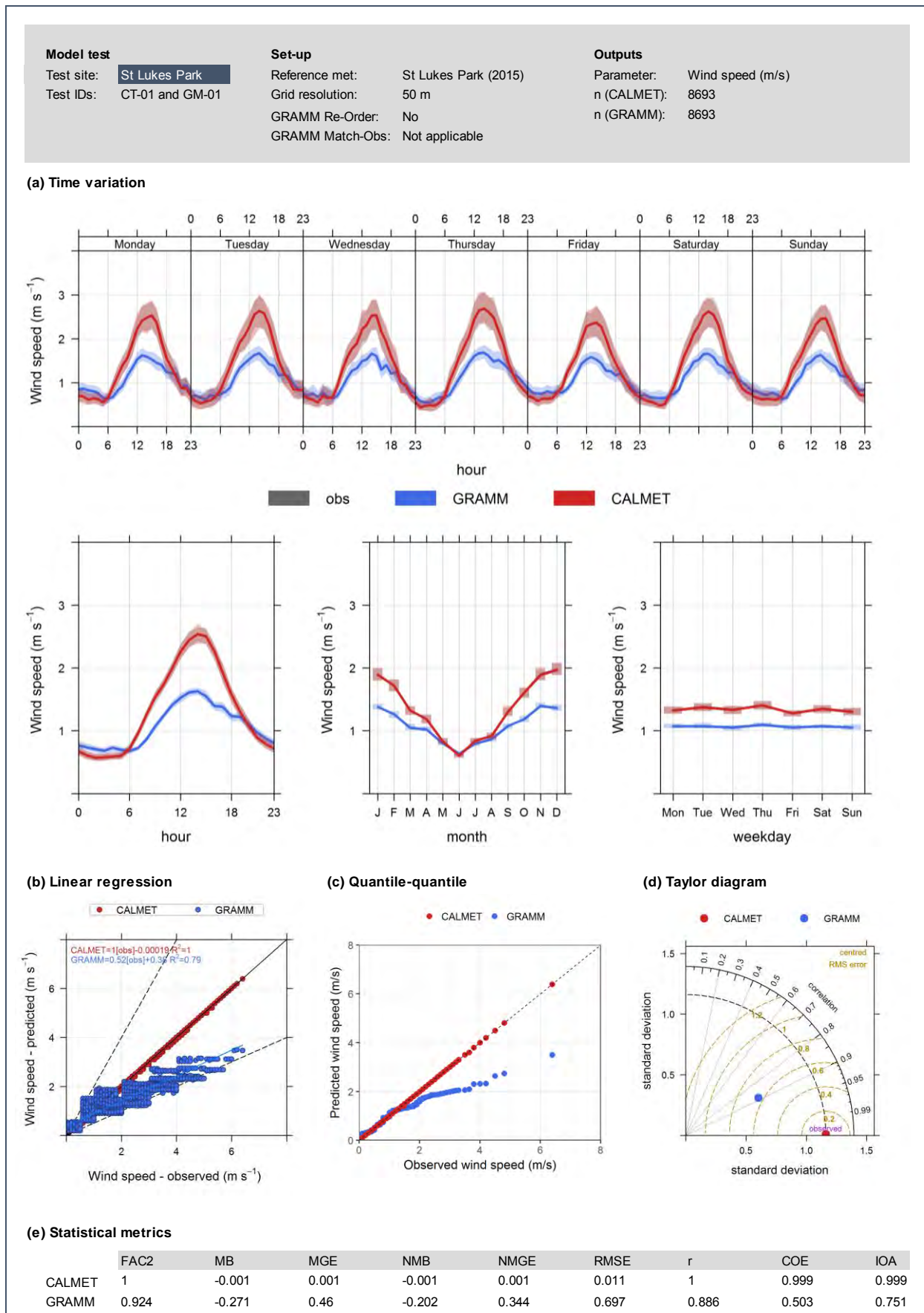


Figure E-2: Meteorological model performance - St Lukes Park extract (tests CT-01 and GM-01)

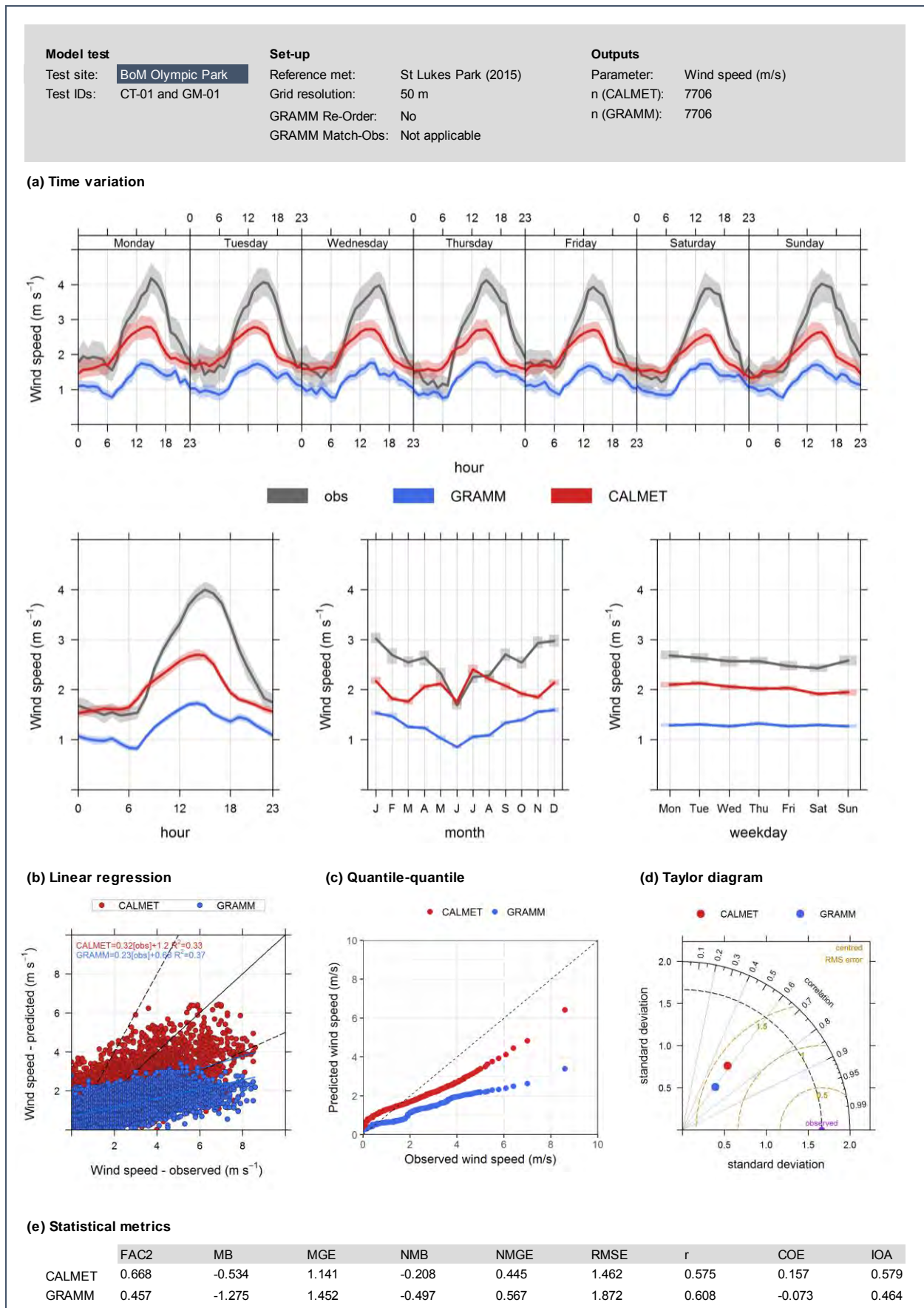


Figure E-3: Meteorological model performance - Sydney Olympic Park extract (tests CT-01 and GM-01)

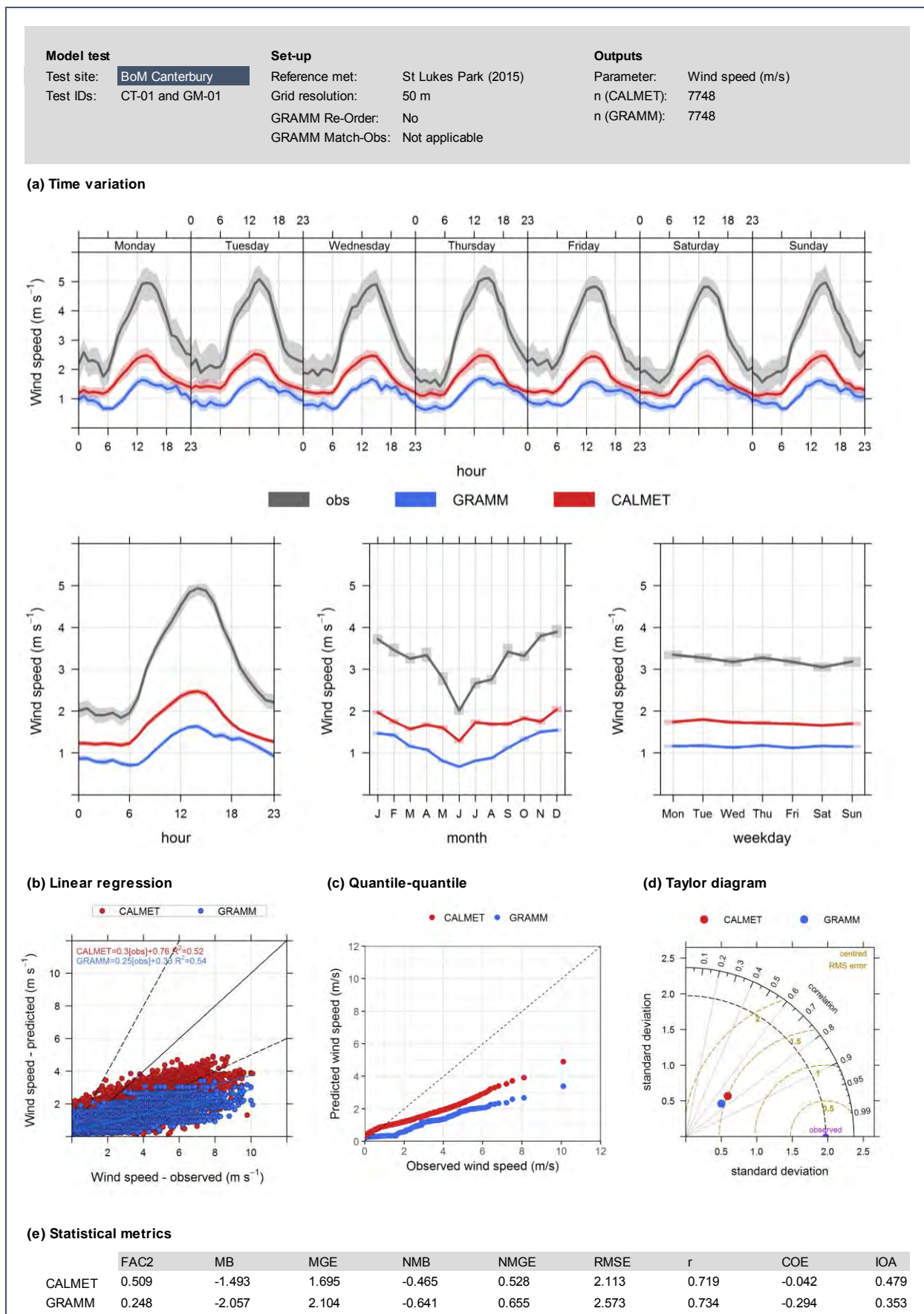


Figure E-4: Meteorological model performance - Canterbury Racecourse extract (tests CT-01 and GM-01)

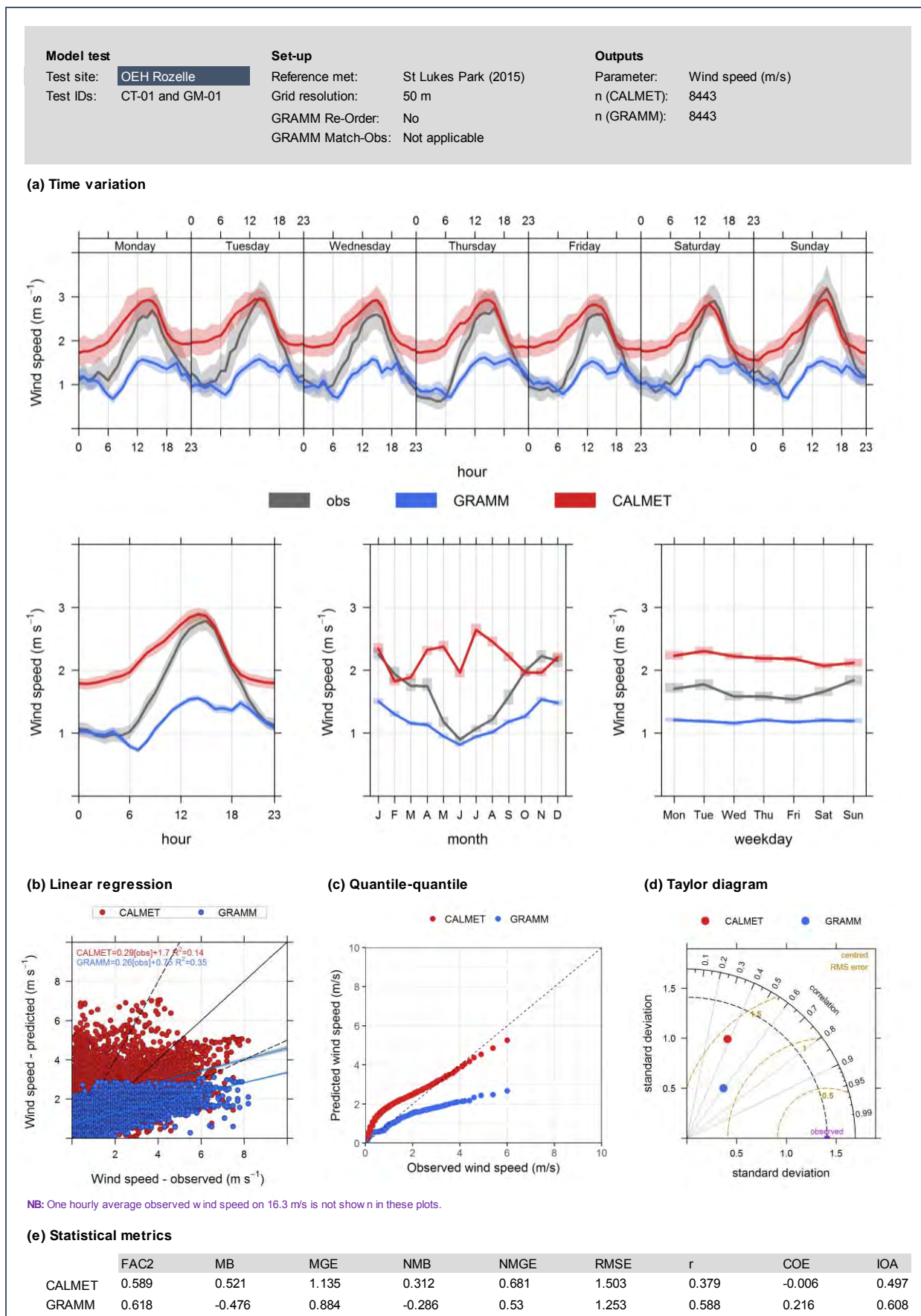


Figure E-5: Meteorological model performance - Rozelle extract (tests CT-01 and GM-01)

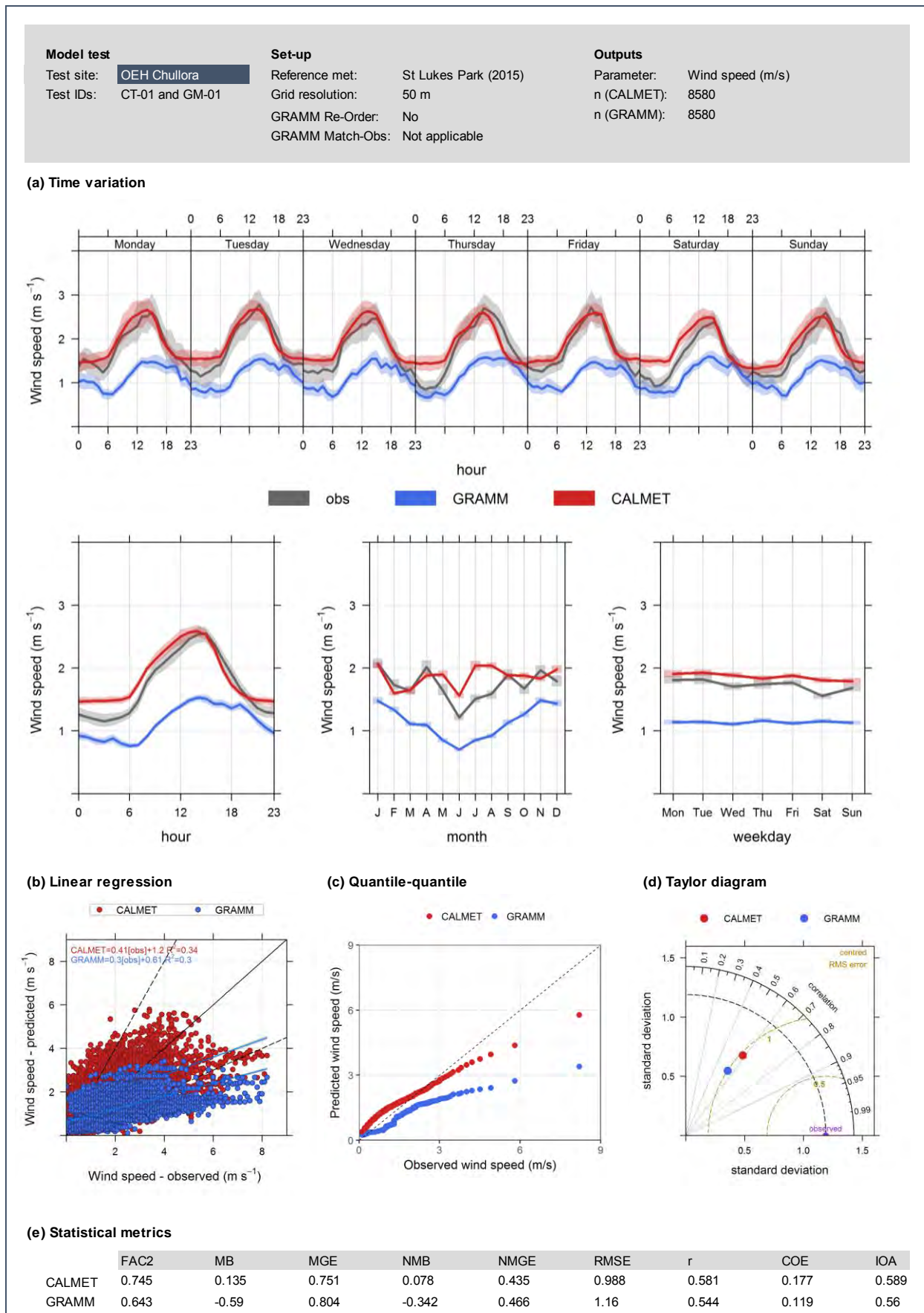


Figure E-6: Meteorological model performance - Chullora extract (tests CT-01 and GM-01)

At Canterbury Racecourse (Figure E-4) CALMET underestimated the wind speed. The underestimation ranged from, on average, around 1 m/s during the night-time to around 2.5 m/s during the mid-afternoon. Again, the underestimation was also greater during the summer than during the winter. The correlation between the predicted and observed values was strong ($R^2 = 0.52$). CALMET overestimated wind speeds where the value for a quantile in the observations was less than around 1 m/s, and underestimated the wind speed for higher quantiles. Again, the GRAMM predictions were systematically lower than those from CALMET (by around 0.6 m/s on average), and showed a poorer temporal agreement with the observations.

In contrast to all other stations, at Rozelle (Figure E-5) CALMET overestimated wind speeds by up to 1 m/s at night-time and gave a more accurate prediction during the afternoon. Moreover, the underestimation was greater during the winter than during the summer. The relationship between the predicted and observed values was weak ($R^2 = 0.14$). GRAMM showed a better agreement with the observations than CALMET at night-time and during winter, and CALMET gave better agreement with observations during the afternoon and during summer. On average, GRAMM underestimated wind speeds by around 0.5 m/s.

At Chullora CALMET performed quite well in terms of the average temporal variation (Figure E-6). The average diurnal variation in wind speed was reproduced well, and the correlation between the hourly paired observations and predictions was moderate ($R^2 = 0.34$). GRAMM again underestimated wind speeds (by around 0.6 m/s on average), and at this station had a weaker correlation than CALMET with observations.

The Taylor diagrams for the stations other than St Lukes Park show that both CALMET and GRAMM predicted wind speeds that had less variation than the observations.

E1.2.1.2 Between-station comparison

The observations, CALMET predictions and GRAMM predictions are compared for all stations in Figure E-7, E-8 and E-9 respectively. These plots show that neither CALMET nor GRAMM predicted the between-station variation in wind speed. In fact, GRAMM predicted little site-to-site variation.

Overall, the results showed that it is a challenge for both CALMET and GRAMM to predict wind speeds accurately across a domain in a situation such as the one investigated, where wind speeds vary quite considerably from location to location.

This is, however, confounded by the use of different instrumentation (e.g. cup-and-vane and sonic anemometers) by BoM and OEH. It is likely that this difference in instrumentation is likely to have played a role in the comparisons.

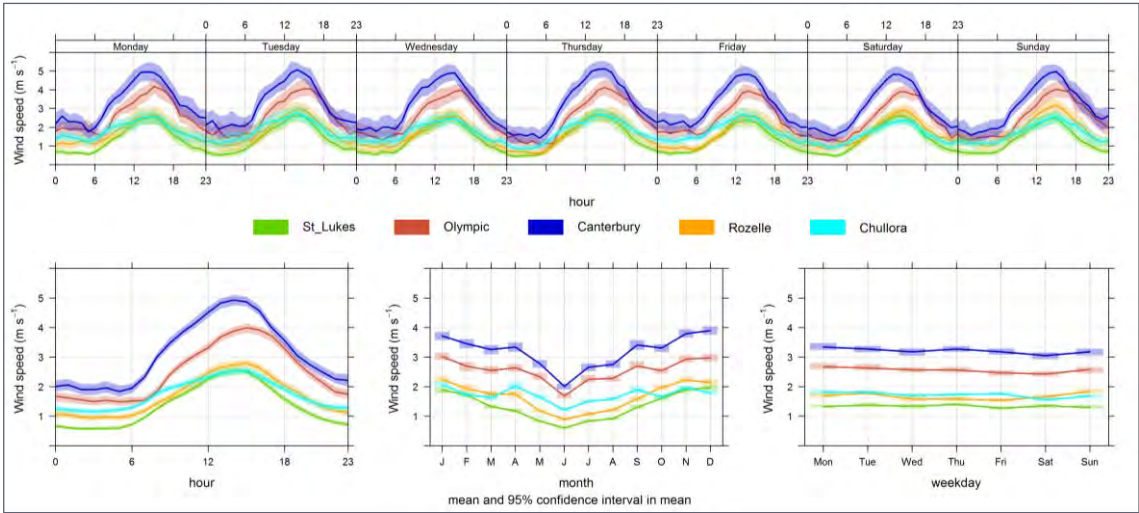


Figure E-7: Time variation for observations at all stations

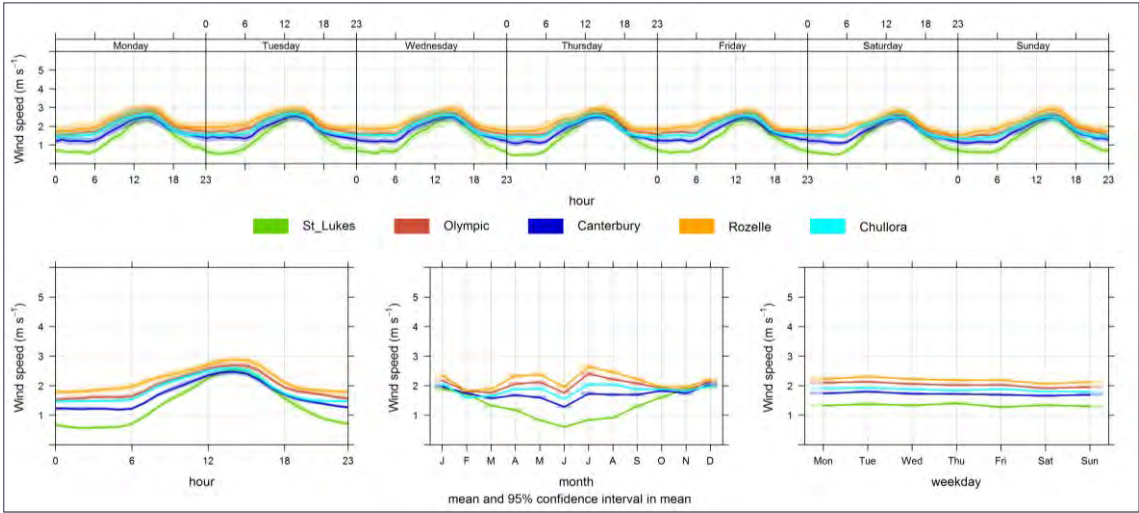


Figure E-8: Time variation for CALMET predictions at all stations

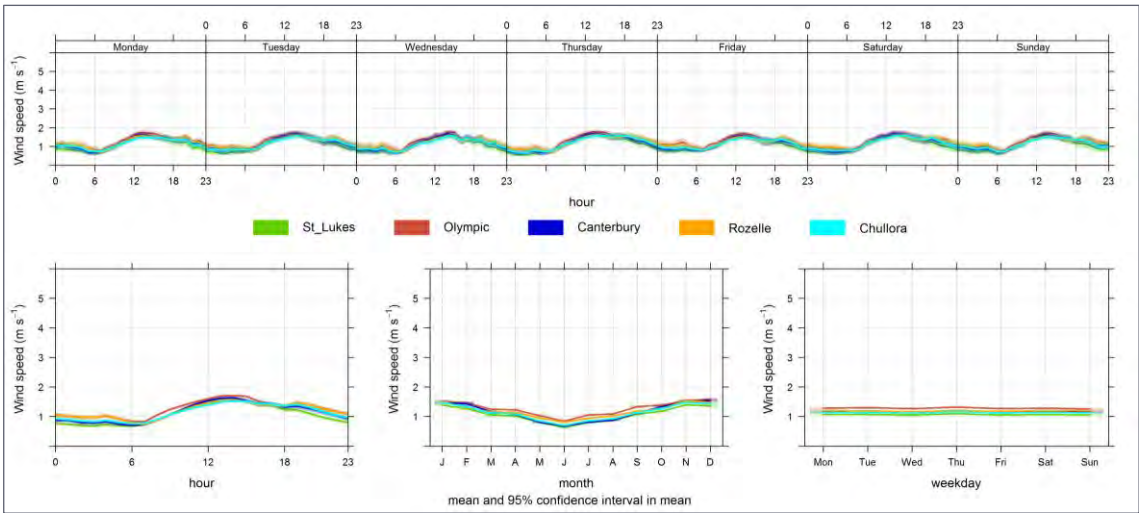


Figure E-9: Time variation for GRAMM predictions at all stations

E1.2.2 GRAMM grid spacing (tests GM-01, GM-02 and GM-03)

Tests GM-01, GM-02 and GM-03 examined the effects of grid spacing within GRAMM. The tests involved running GRAMM with reference meteorology taken from the St Lukes Park station only, and with grid spacing of 50 metres, 100 metres and 200 metres respectively. CALMET was not included in these tests.

The results of the tests for the various monitoring stations are shown in Figure E-10 to E-14.

Notwithstanding the general under-prediction by GRAMM identified earlier, the use of a 50 metre grid generally resulted in better predictions than the 100 metre and 200 metre grids. However, the results for the different grid spacings were generally quite similar. This implies that for a simulation of this kind (flat terrain, no buildings, single station reference meteorology), the results will not be very dependent on the GRAMM grid resolution. In other words, the effect of grid resolution is likely to be small relative to the differences between the predictions and the observations.

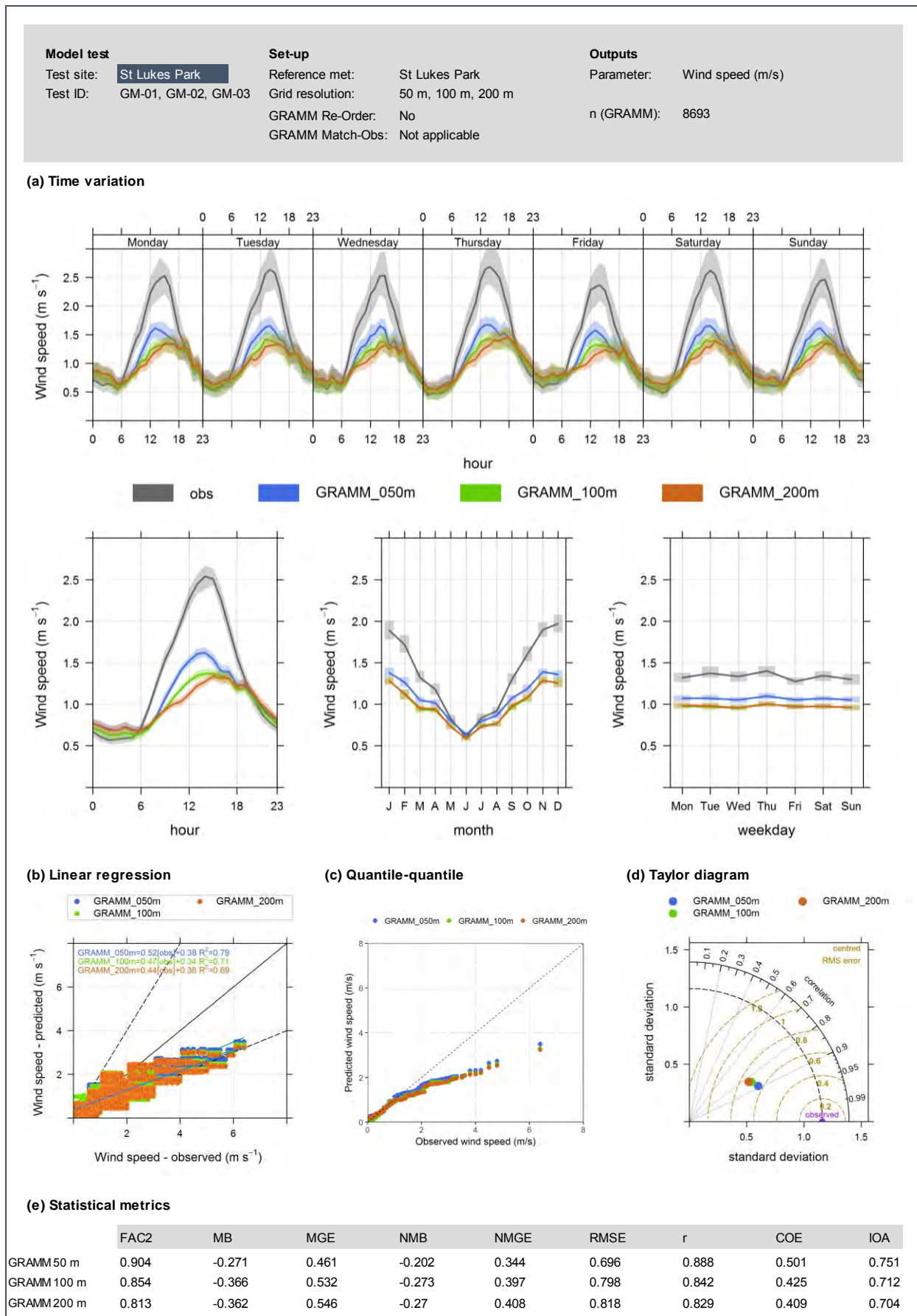


Figure E-10: Effects of grid spacing in GRAMM – St Lukes Park extract (tests GM-01, GM-02, and GM-03)

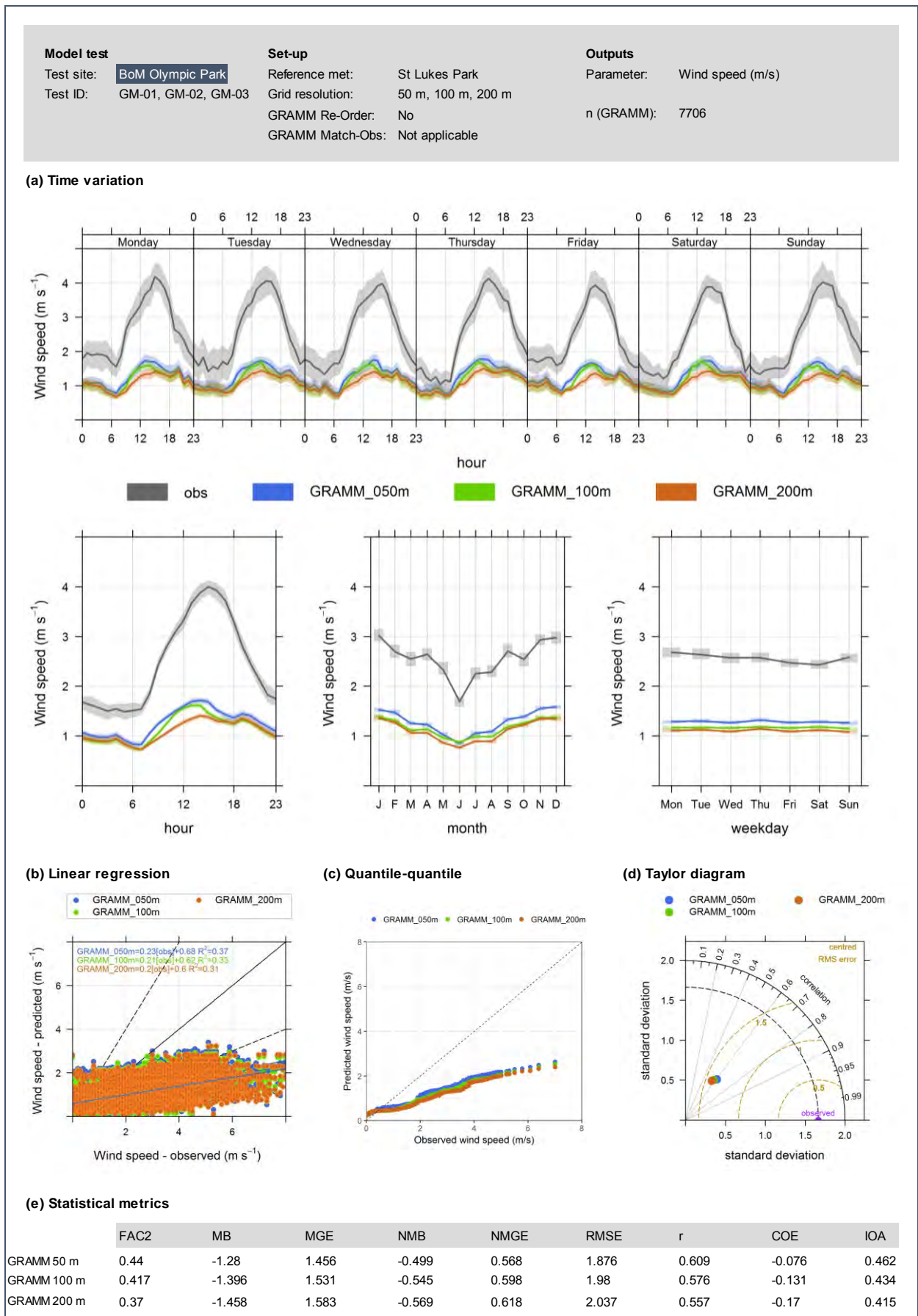


Figure E-11: Effects of grid spacing in GRAMM – Sydney Olympic Park extract (tests GM-01, GM-02, and GM-03)

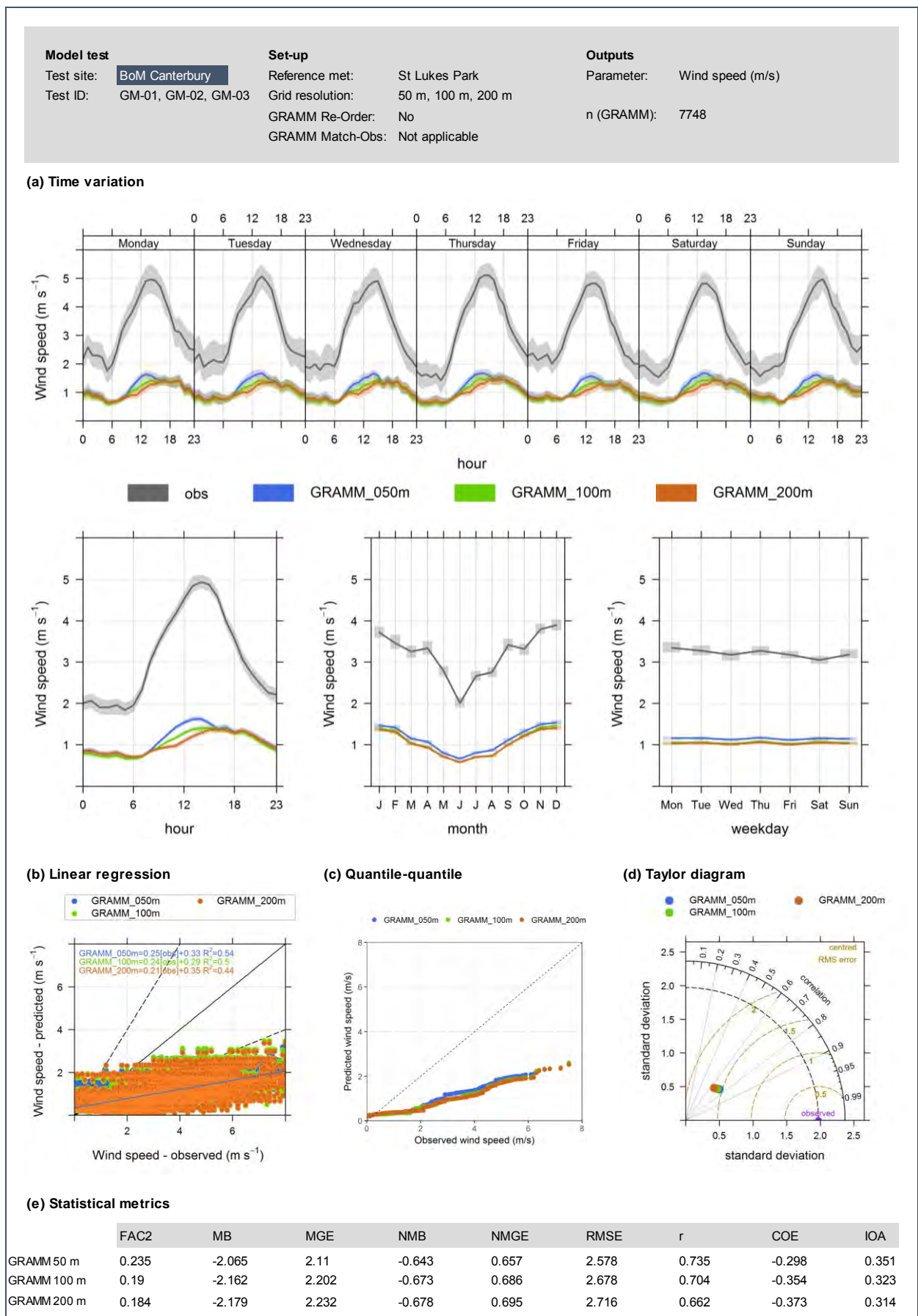


Figure E-12: Effects of grid spacing in GRAMM – Canterbury Racecourse extract (tests GM-01, GM-02, and GM-03)

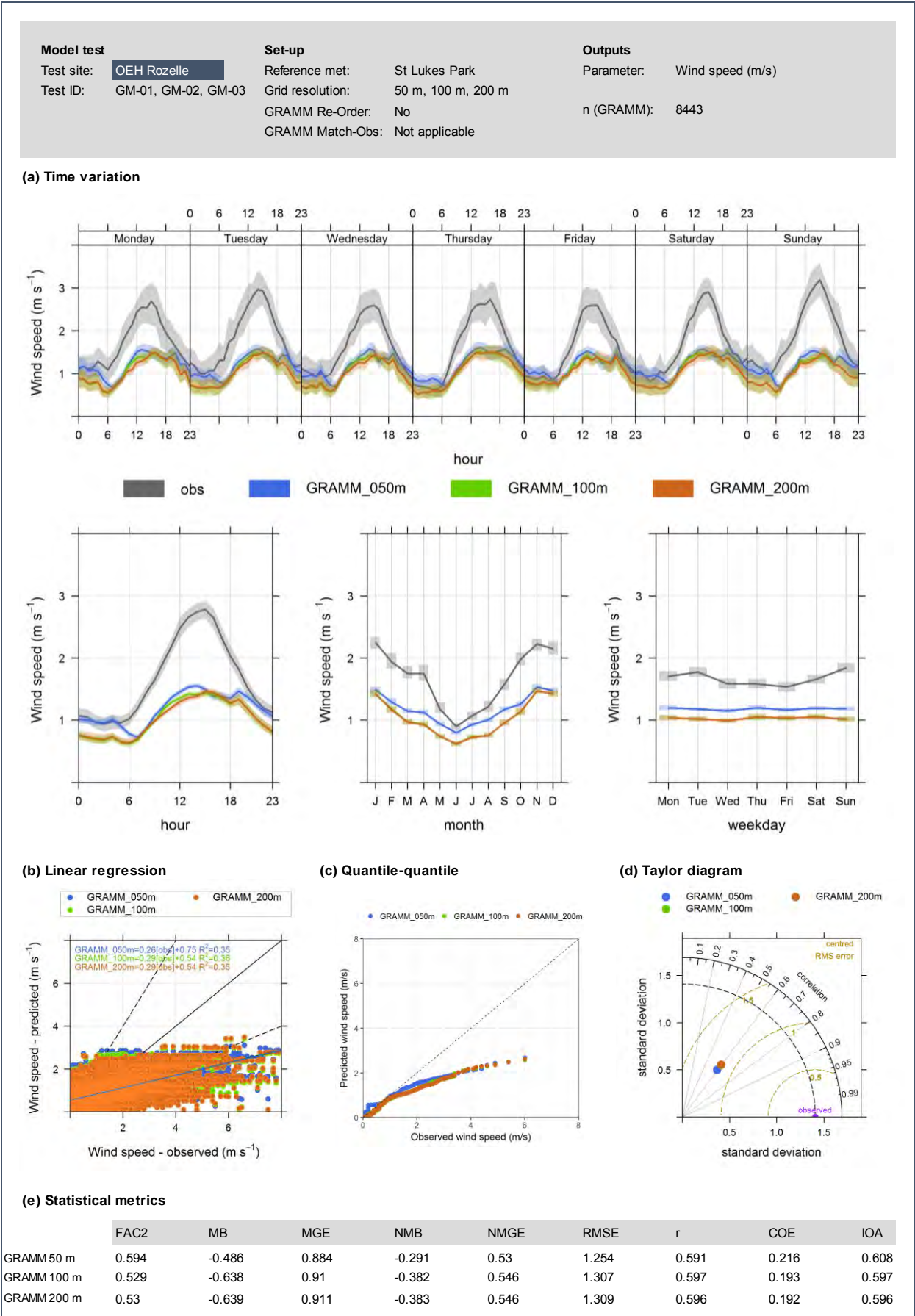


Figure E-13: Effects of grid spacing in GRAMM – Rozelle extract (tests GM-01, GM-02, and GM-03)

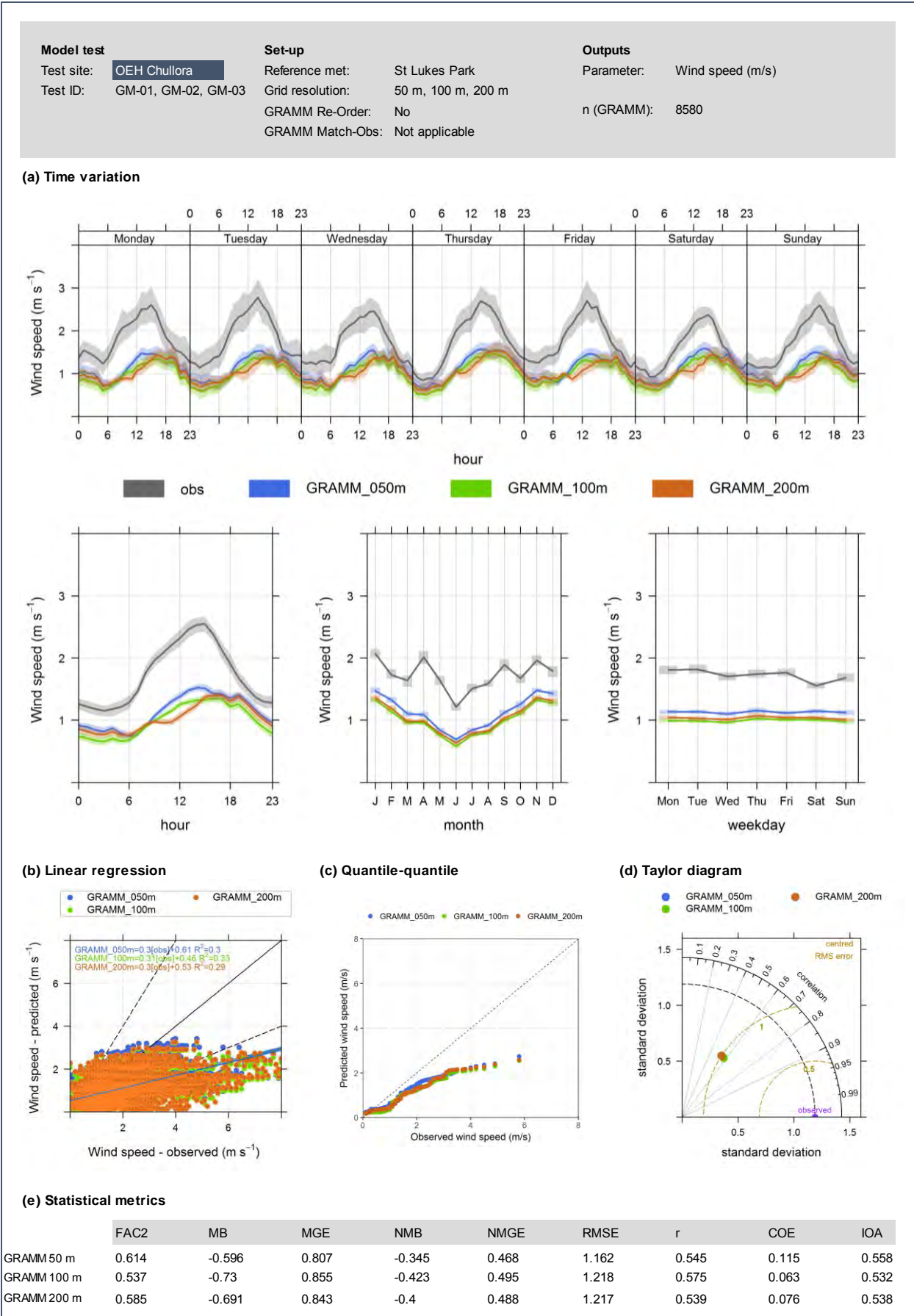


Figure E-14: Effects of grid spacing in GRAMM – Chullora extract (tests GM-01, GM-02, and GM-03)

E1.2.3 GRAMM Re-Order (tests GM-01 and GM-04)

Tests GM-01 and GM-04 both involved running GRAMM with reference meteorology taken from the St Lukes Park station only, and with grid spacing of 50 metres. A comparison between these two tests illustrated the effects of running GRAMM without the Re-Order function (test GM-01) and with the Re-Order function (test GM-04). The results of the tests are shown for all monitoring stations in Figures E-15 to E-19, along with the observations and the results from CALMET test CT-01.

The use of the Re-Order function resulted in a general improvement in the GRAMM predictions. Unsurprisingly, this improvement was most pronounced at the St Lukes Park station, where the R^2 value increased from 0.79 to 0.88. At this location the average wind speed predicted by GRAMM increased by only around 0.1 m/s, but there was a noticeably better prediction of higher wind speeds.

Although there was also an improvement in the GRAMM predictions at the other stations, this was less pronounced. The main effects of the Re-Order function were as follows:

- It slightly improved the prediction of the higher wind speeds. Conversely, the prediction of low wind speeds was made slightly worse.
- It gave a slight improvement in the correlation between the predictions and the observations.
- It resulted in a variability in the predictions that was slightly closer to that in the observations.

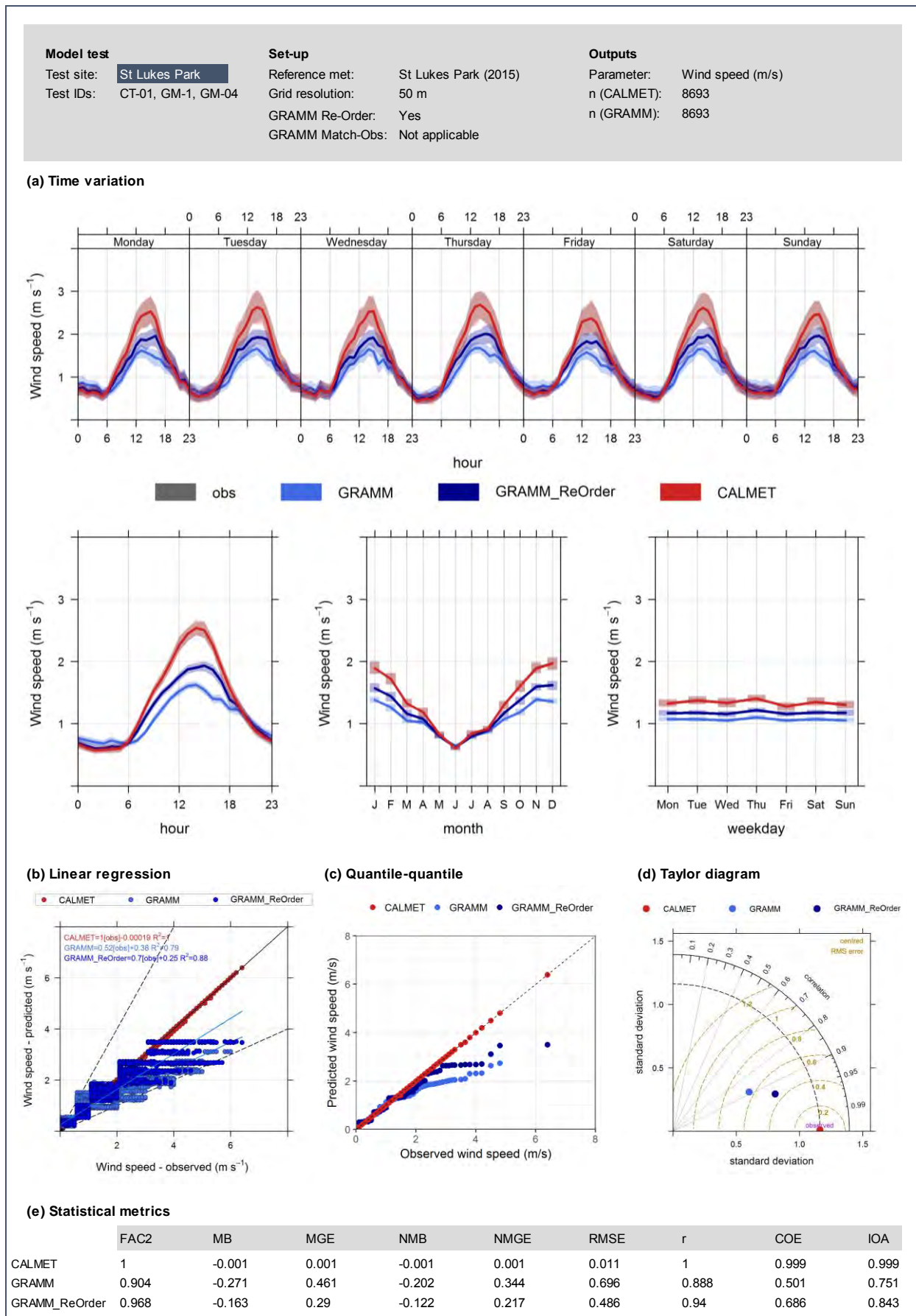


Figure E-15: Effects of Re-Order in GRAMM – St Lukes Park extract (tests CT-01, GM-01, and GM-04)

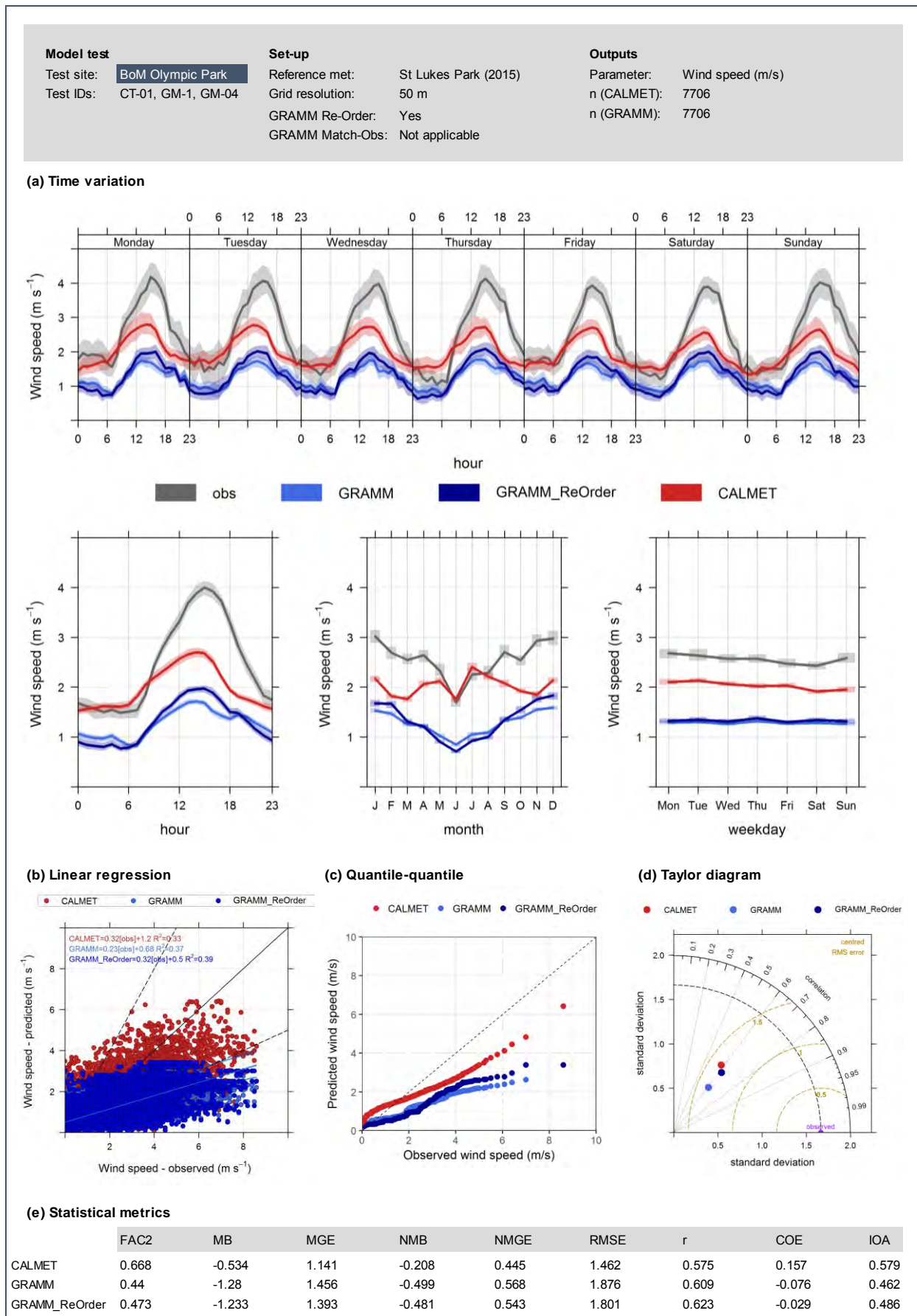


Figure E-16: Effects of Re-Order in GRAMM – Sydney Olympic Park extract (tests CT-01, GM-01, and GM-04)

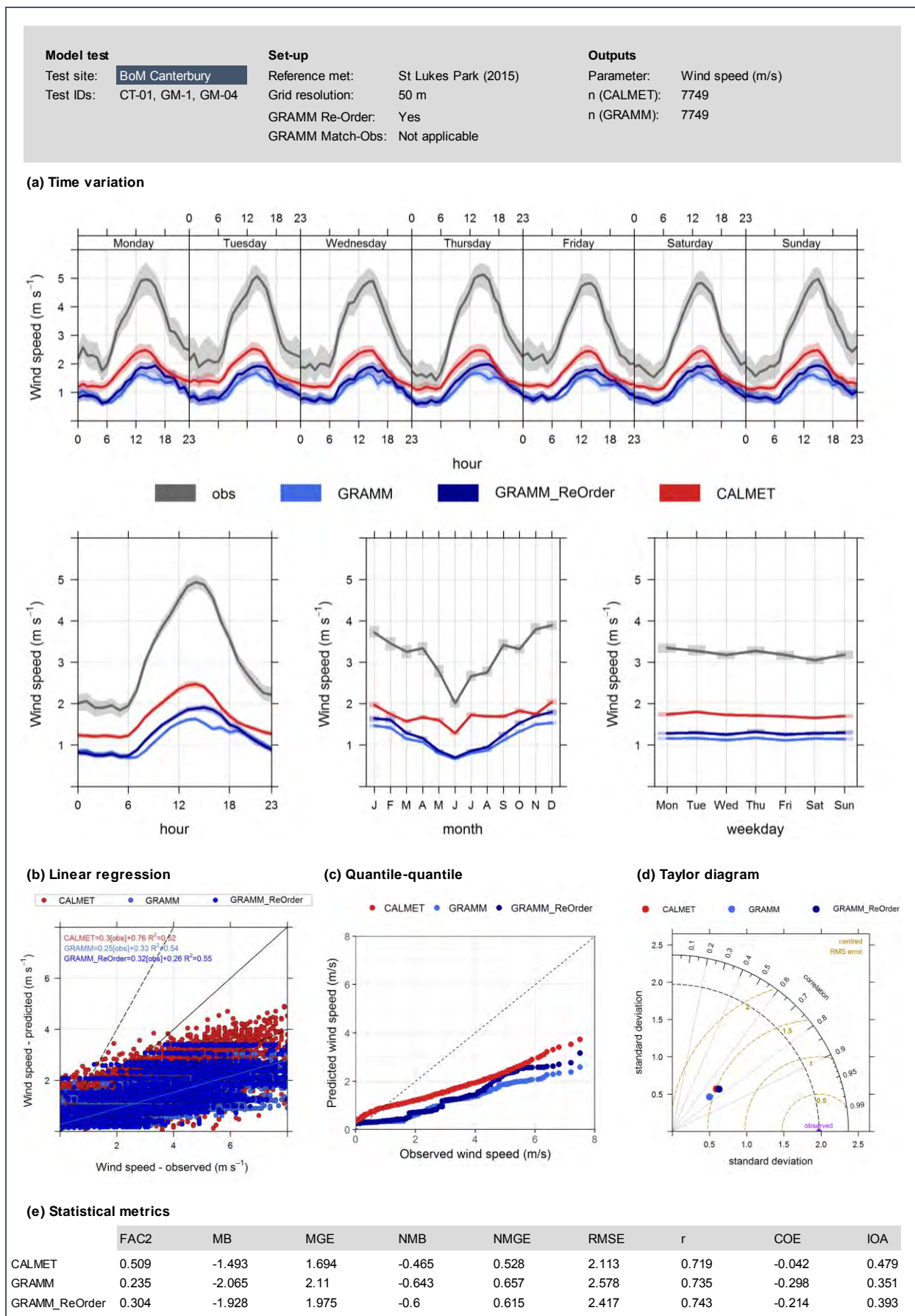


Figure E-17: Effects of Re-Order in GRAMM – Canterbury Racecourse extract (tests CT-01, GM-01, and GM-04)

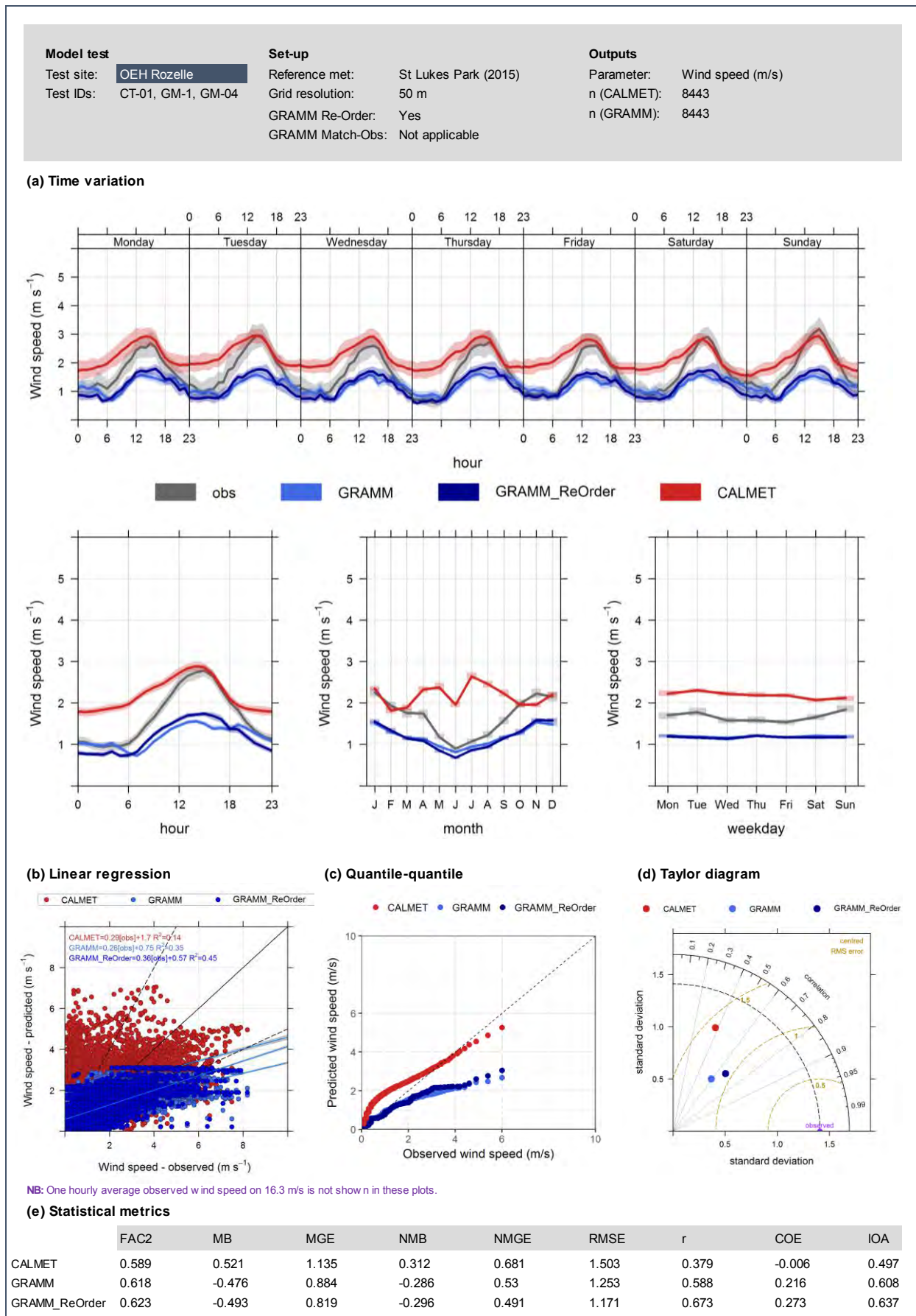


Figure E-18: Effects of Re-Order in GRAMM – Rozelle extract (tests CT-01, GM-01, and GM-04)

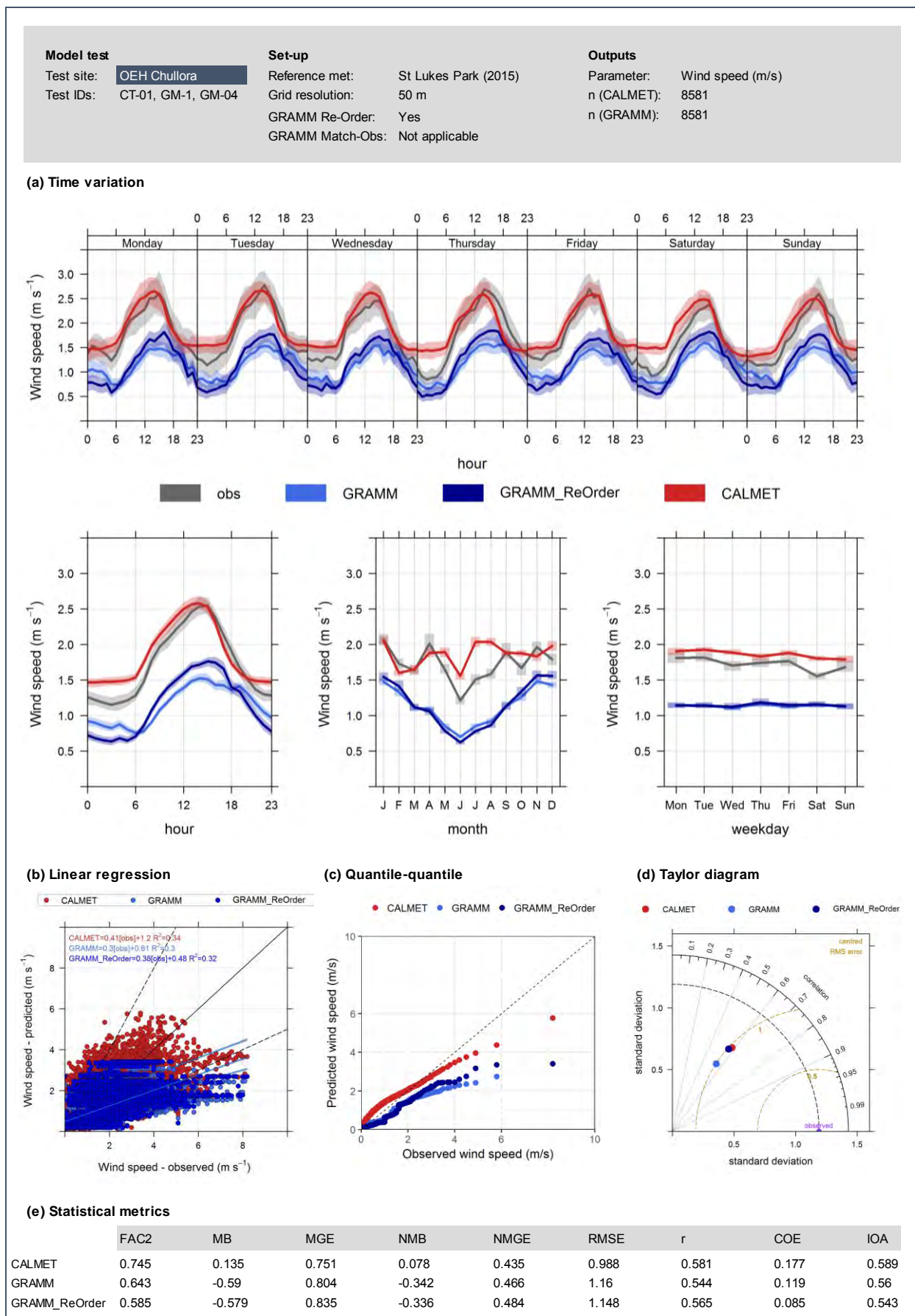


Figure E-19: Effects of Re-Order in GRAMM – Chullora extract (tests CT-01, GM-01, and GM-04)

E1.3 Series B (multiple station or synthetic reference meteorology)

E1.3.1 CALMET vs GRAMM (tests CT-02 and GM-05)

In contrast to tests CT-01 and GM-01, where the only reference meteorology was St Lukes Park, these tests illustrated the behaviour of CALMET when several meteorology stations were used as input, except the station for which CALMET predictions were obtained (in this case, St Lukes Park). This provided a more independent test of model performance at St Lukes Park than tests CT-01 and GM-01.

In Figure E-20 for the St Lukes Park extract, it can be seen that the performance of CALMET deteriorated significantly ($R^2 = 0.43$) compared with the 'perfect' performance in test CT-01. With reference meteorology not including this station, the performance of CALMET was similar to that of GRAMM. The following points can also be noted:

- Wind speeds were systematically overestimated by both CALMET and GRAMM, and low wind speeds were overestimated more than high wind speeds. To some extent this would have been a consequence of the measured wind speeds at St Lukes Park, which were the lowest of all the stations.
- On average, the performance of GRAMM was better than that of CALMET at low wind speeds. CALMET gave slightly better predictions at high wind speeds.
- Both models gave a variability in predictions that was quite close to that in the measurements (CALMET was slightly better than GRAMM).
- On an hourly basis the performance of both models was strong ($R^2 = 0.43$ to 0.47).

At the other stations (Figures E-21 to E-24) the performance of GRAMM in series B was markedly better than in series A. For example, when comparing test GM-05 with test GM-04 (with Re-Order), R^2 values increased from 0.32-0.55 to 0.49-0.70), and the proportion of predictions within a factor of two of observations increase from 0.30-0.62 to 0.64-0.78

As the other stations were included in CALMET, the expected 'perfect' performance was observed.

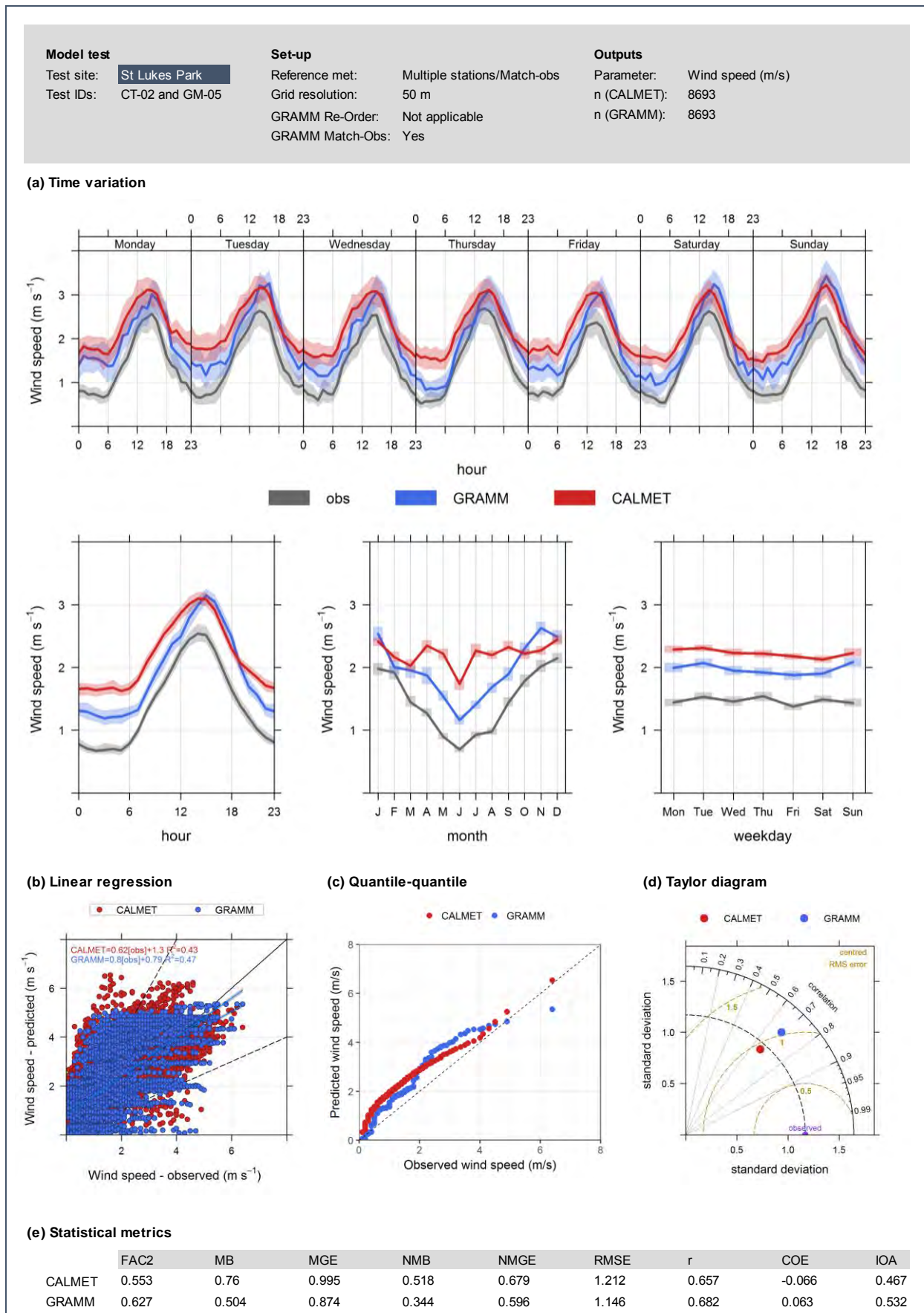
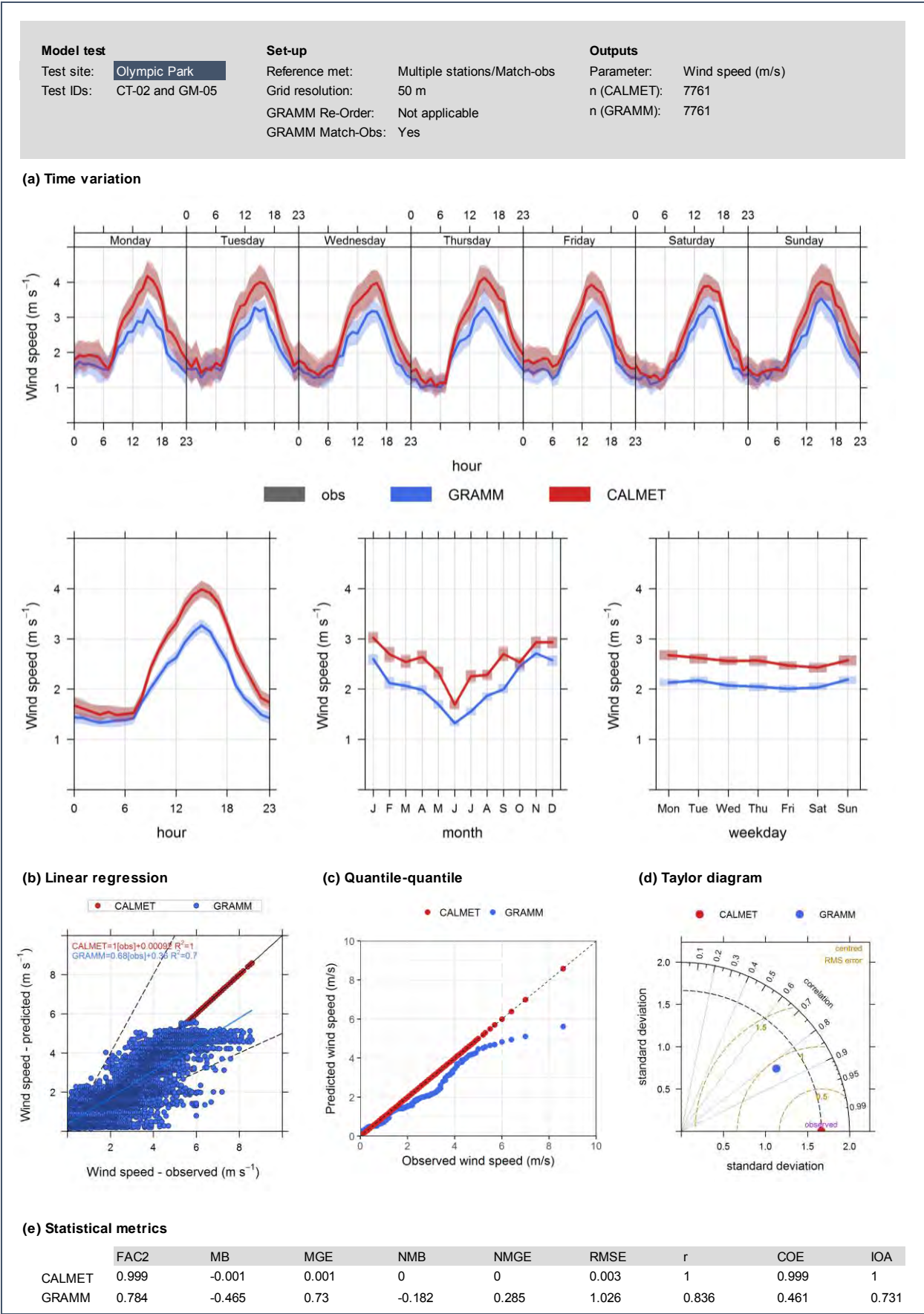


Figure E-20: Meteorological model performance - St Lukes Park extract (tests CT-02 and GM-05)



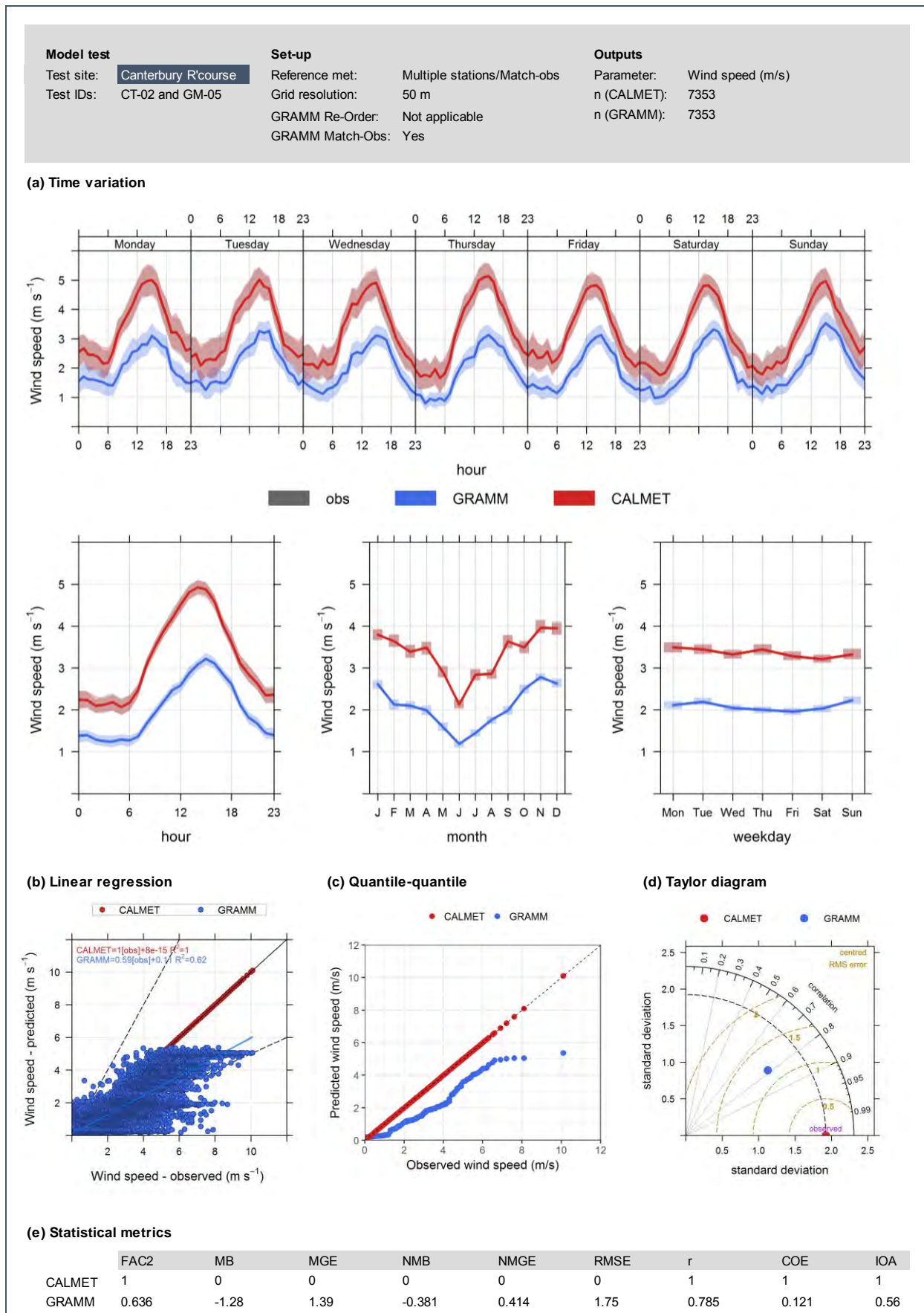


Figure E-22: Meteorological model performance - Canterbury Racecourse extract (tests CT-02 and GM-05)

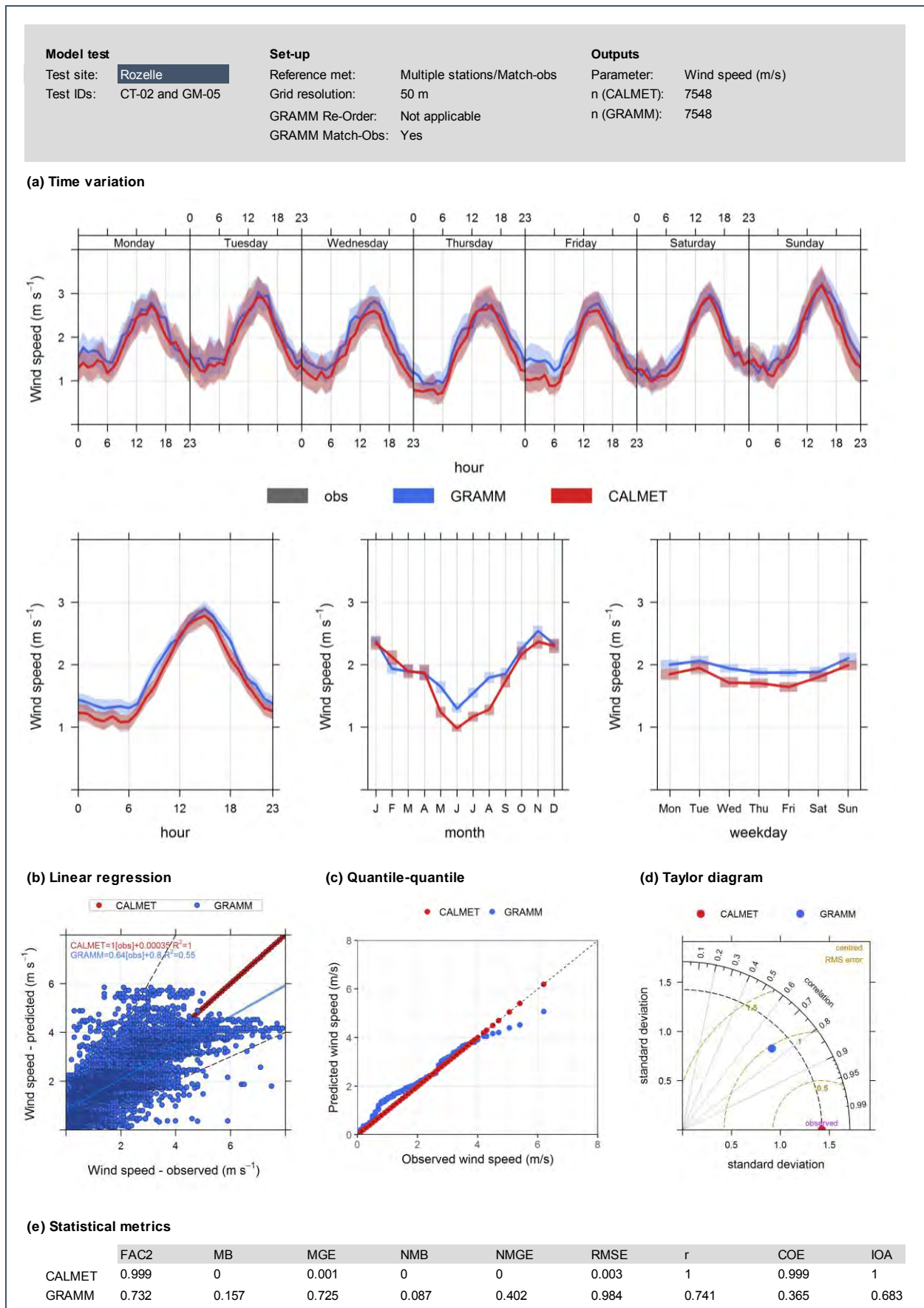


Figure E-23: Meteorological model performance – Rozelle extract (tests CT-02 and GM-05)

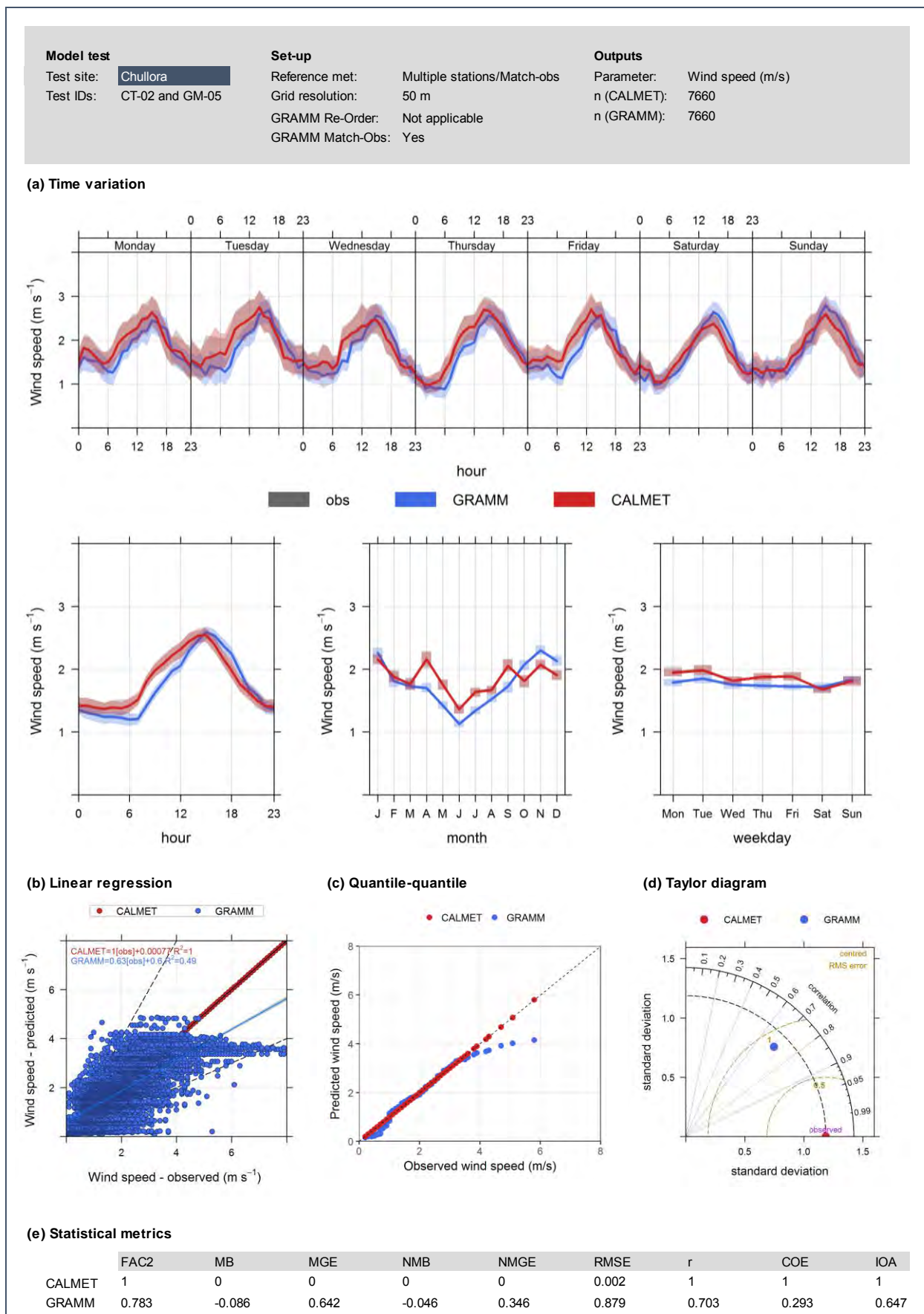


Figure E-24: Meteorological model performance – Chullora extract (test CT-02 and GM-05)

E1.3.2 CALMET vs GRAMM (tests CT-02 and GM-06)

In these tests the Match-to-Observations function in GRAMM was applied to all stations, and the results are shown in Figures E-25 to E-29. No changes to CALMET test CT-02 were made. The CALMET extracts for the various stations exhibited the 'perfect' as before, and are just included here for completeness. The CALMET results for St Lukes Park are also shown. As St Lukes Park was not included in the reference meteorology for test CT-02, the 'perfect' results for test CT-01 are shown for this station. It is the effects of GRAMM that are of more interest here.

For St Lukes Park the performance of GRAMM was much improved relative to tests GM-01, GM-04 and GM-05. It should be noted that the ability of GRAMM to simulate the wind speeds at St Lukes Park was still somewhat constrained by the algorithms in the Match-to-Observations function. This is because the function provides an optimised fit across all reference stations included, and therefore for some meteorological situations the fit will have been poorer at St Lukes Park than at the other stations. In a model domain where meteorological data at multiple stations is similar (and often expected particularly in flat terrain conditions), this would be less of an issue. However, as previously explained, there is some significant variation in collected wind data across the model domain due to different in meteorological instruments and specific station location etc. It is important to note that (and as seen in the previous section's CALMET results, that CALMET will only perform a 'perfect' match at the location at which the meteorological data was entered. At locations where there is no input meteorological data, the match is much poorer.

For the other stations the results for test GM-06 were actually worse than in test GM-05, with the exception of Rozelle. It therefore appears that simply adding more reference stations to GRAMM does not automatically improve its performance likely due to the reasons stated above regarding significant difference in meteorological input data within the model domain.

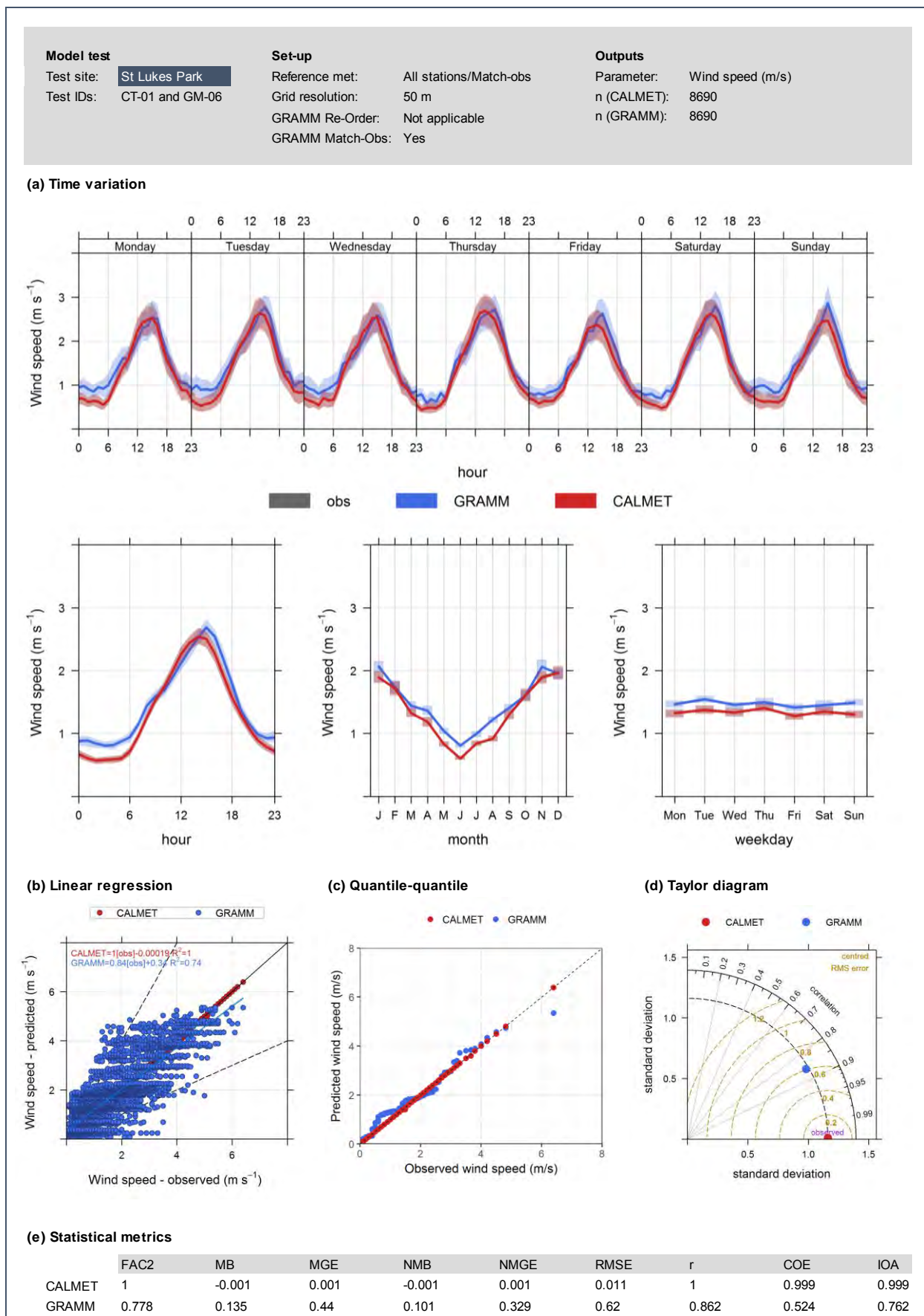


Figure E-25: Meteorological model performance – St Lukes Park extract (tests CT-01 and GM-06)

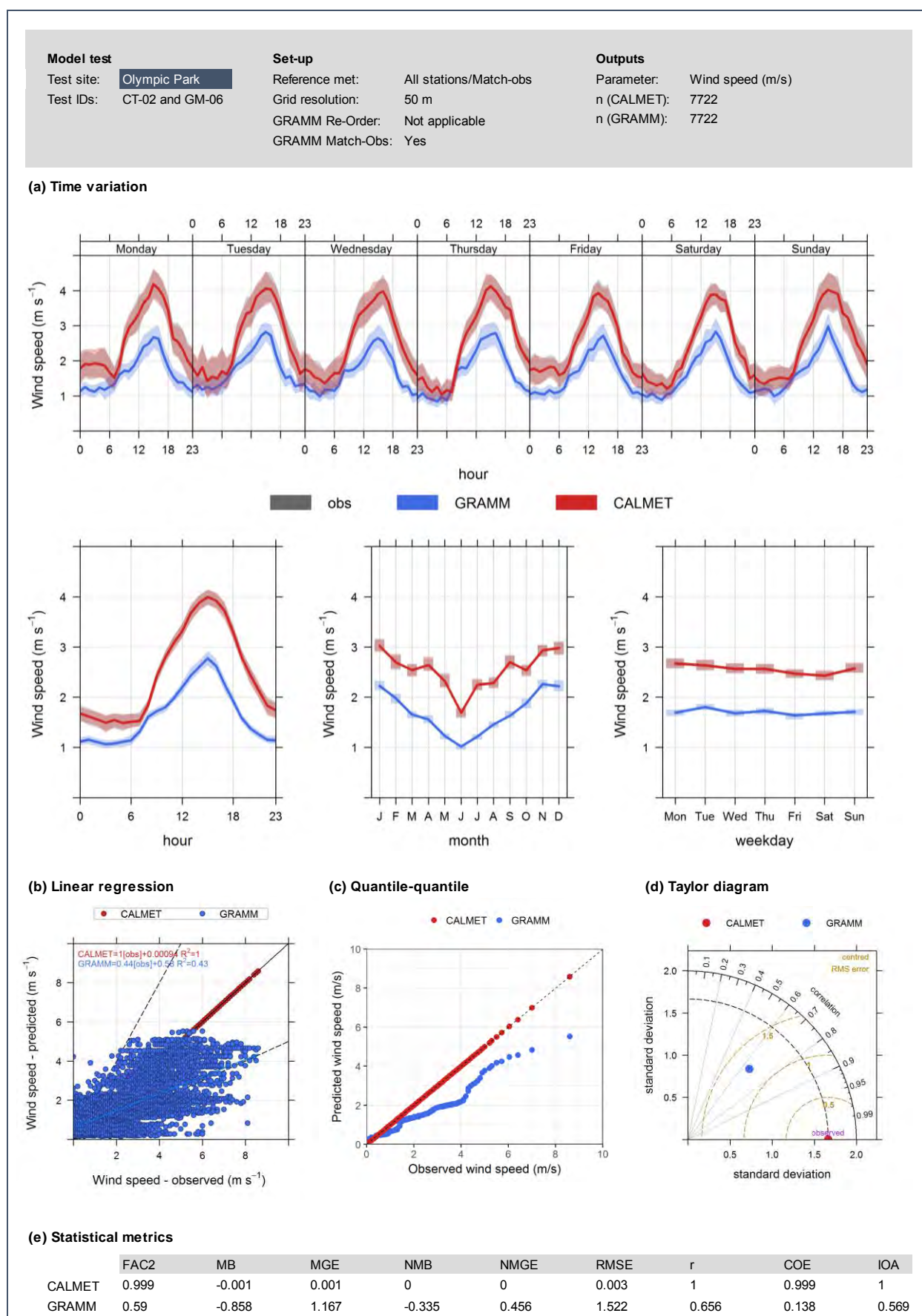


Figure E-26: Meteorological model performance – Sydney Olympic Park extract (tests CT-02 and GM-06)

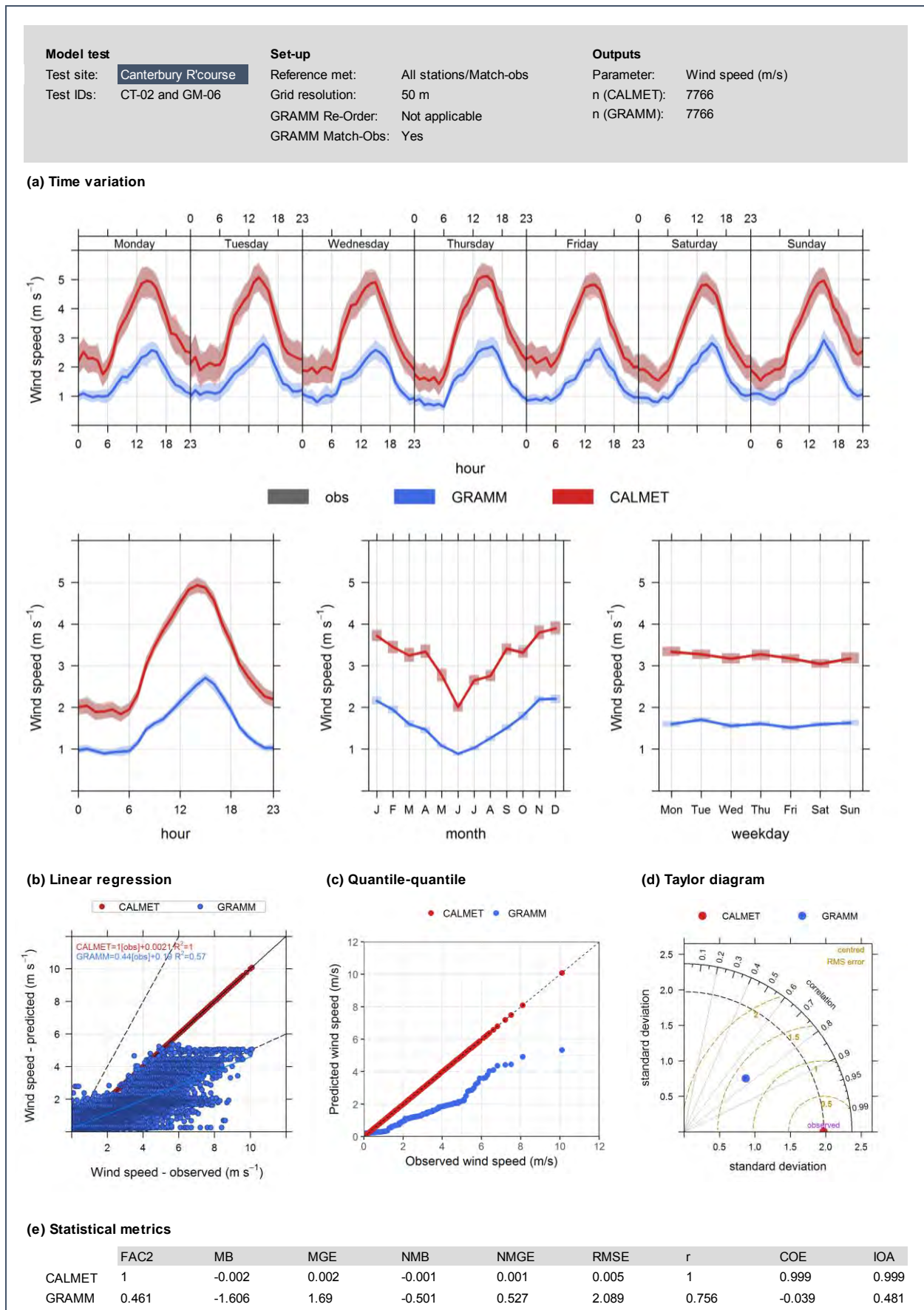


Figure E-27: Meteorological model performance – Canterbury Racecourse extract (tests CT-02 and GM-06)

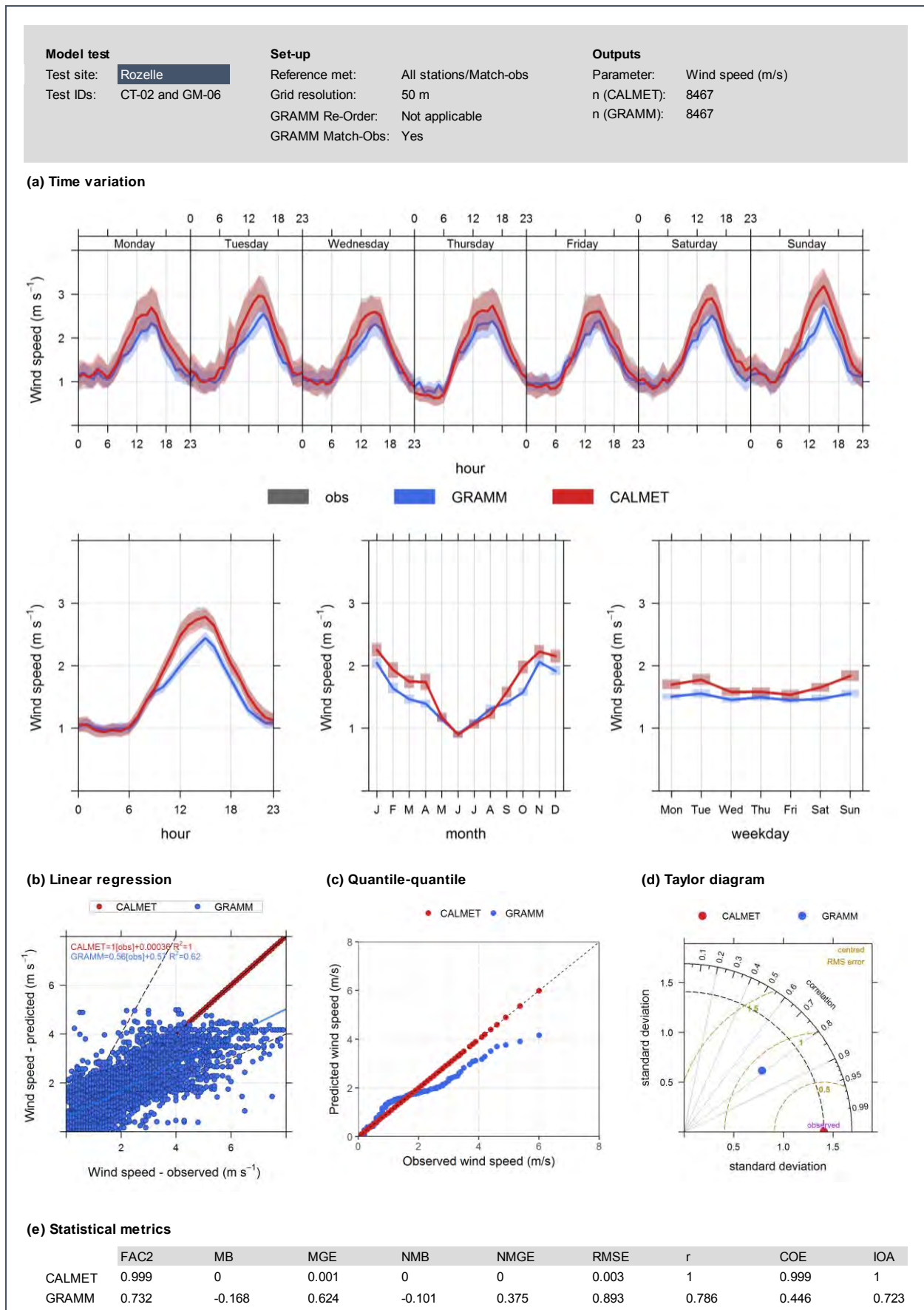


Figure E-28: Meteorological model performance – Rozelle extract (tests CT-02 and GM-06)

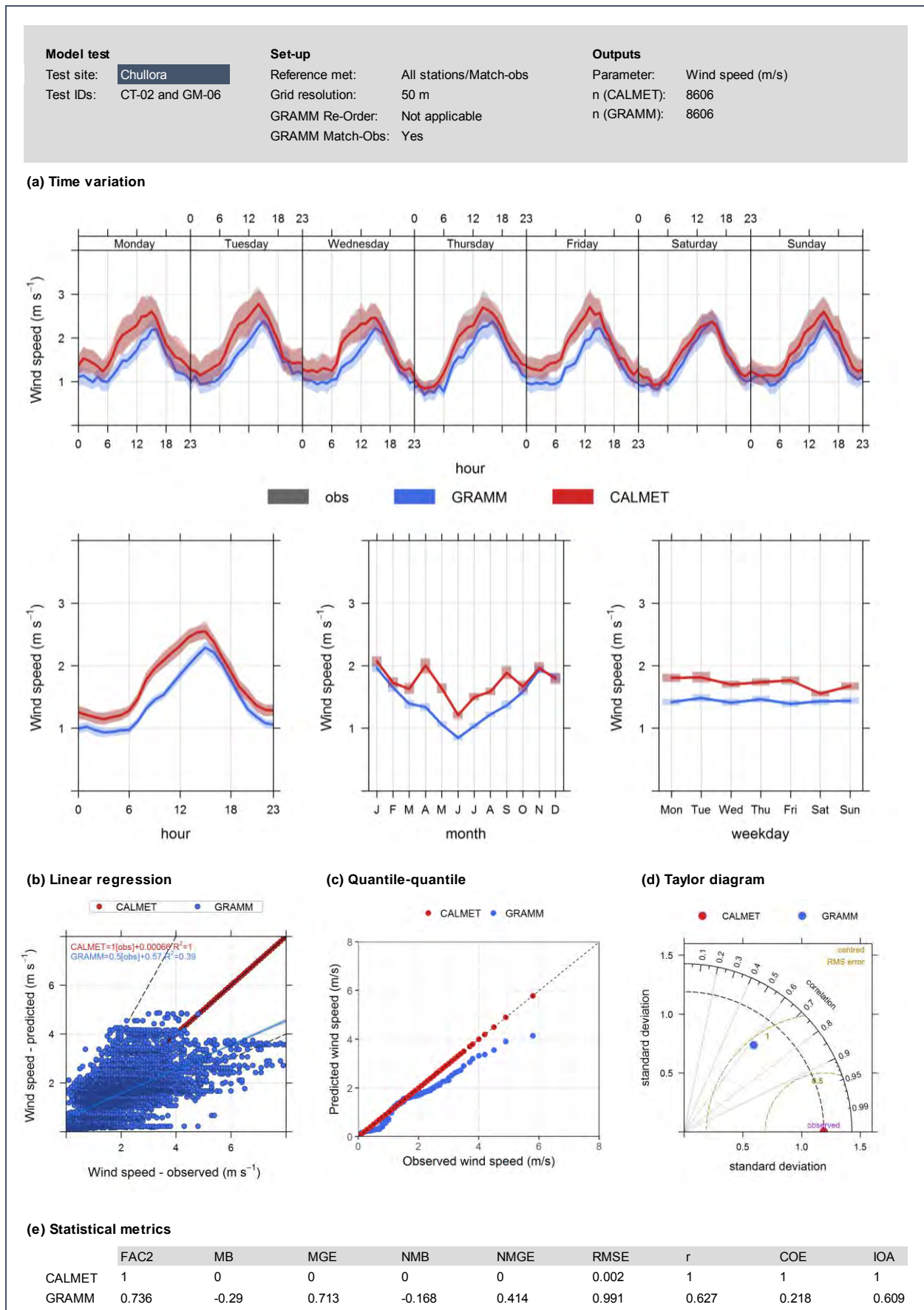


Figure E-29: Meteorological model performance – Chullora extract (tests CT-02 and GM-06)

E1.3.3 GRAMM Match-to-Observations at single station (tests CT-01 and GM-07)

In these tests GRAMM was run with Match-to-Observations for St Lukes Park only, and the predictions were compared with those from CALMET test CT-01 from series A. The results are presented in Figures E-30 to E-34.

Given that GRAMM was not constrained by the need to match the observations at other stations, this test gave by far the best performance of any GRAMM tests for St Lukes Park ($R^2 = 0.96$).

At the other stations neither GRAMM nor CALMET gave systematically the better performance. The results for GRAMM were similar to those from test GM-06.

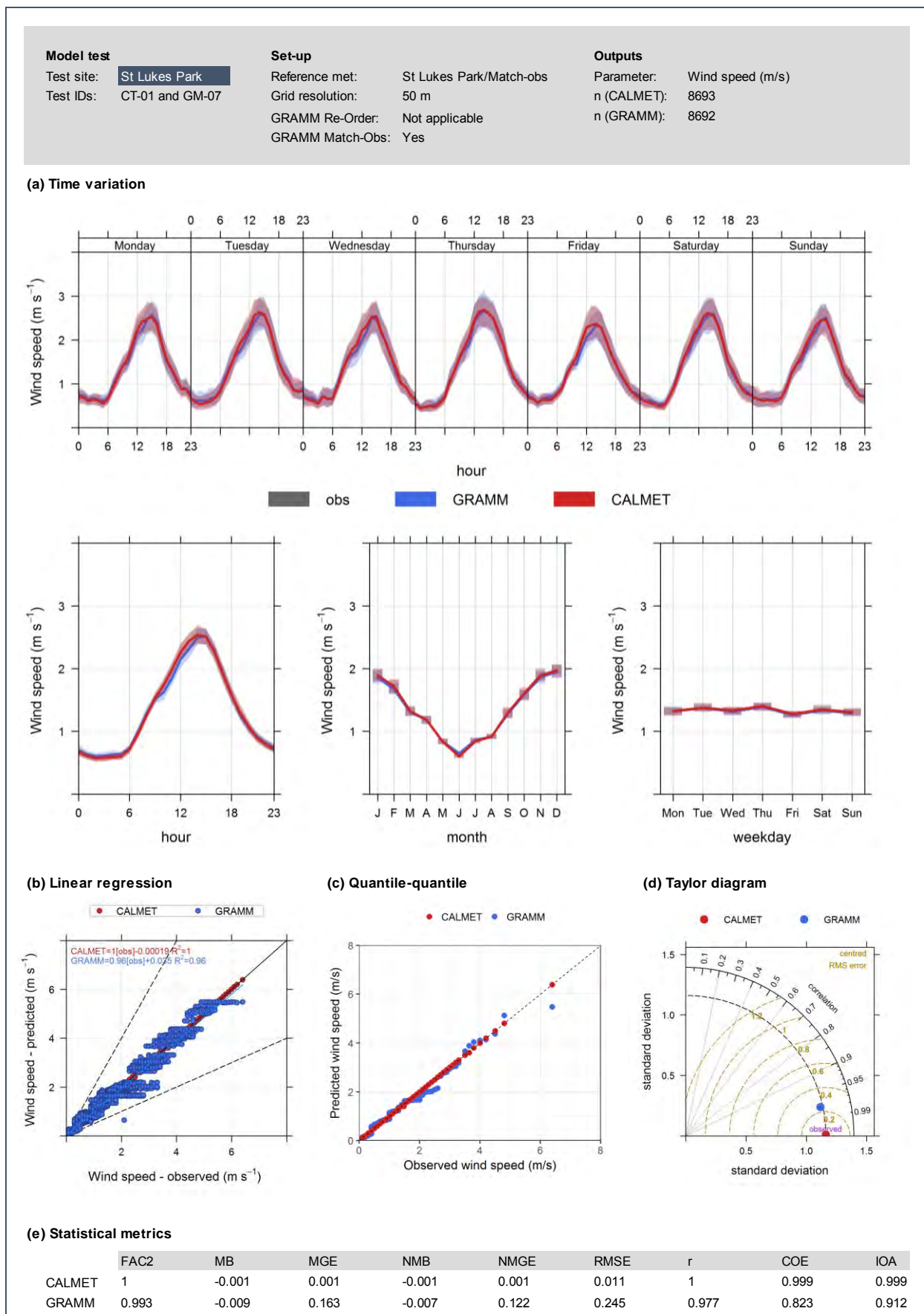


Figure E-30: Meteorological model performance – St Lukes Park extract (tests CT-01 and GM-07)

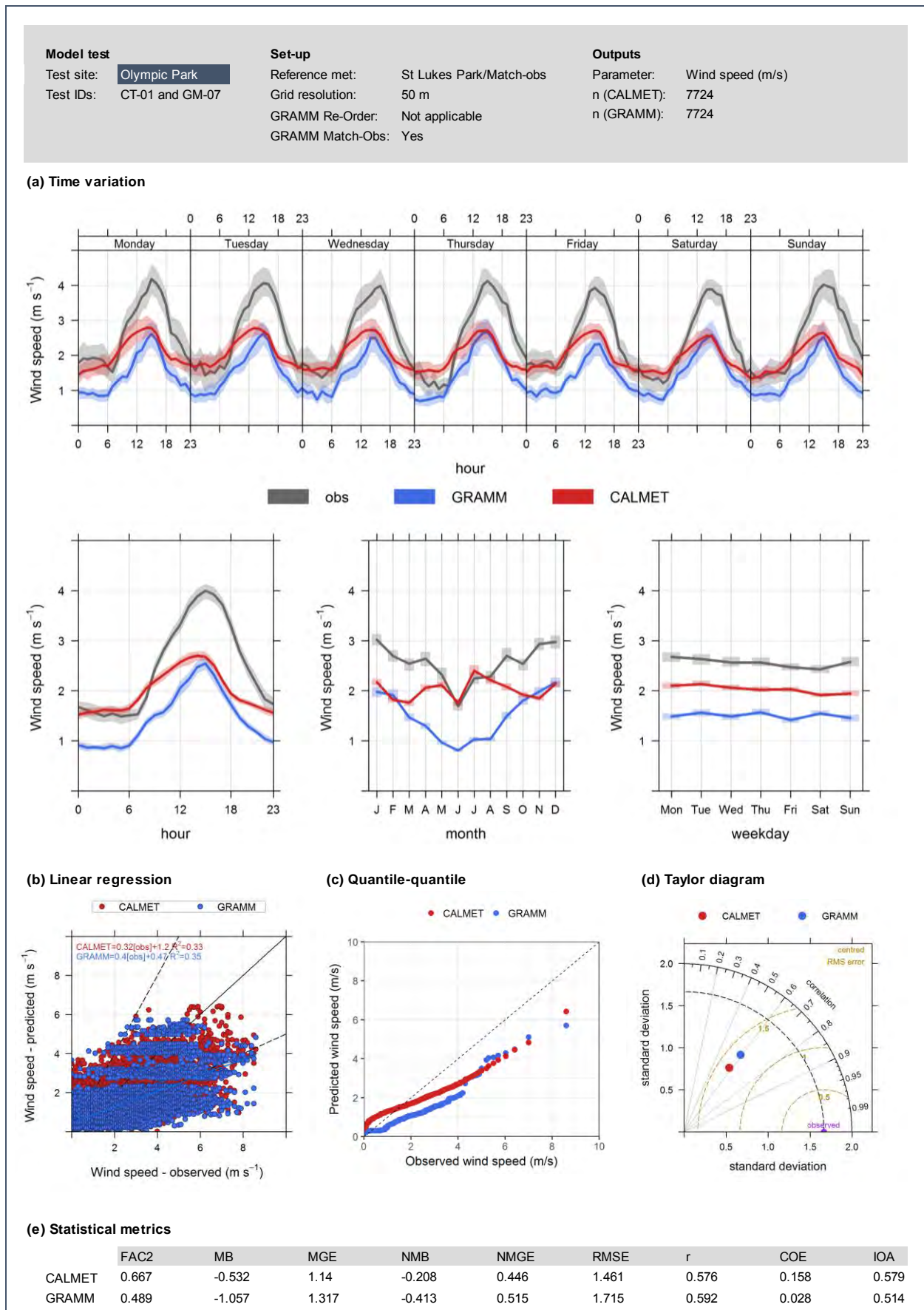


Figure E-31: Meteorological model performance – Olympic Park extract (tests CT-01 and GM-07)

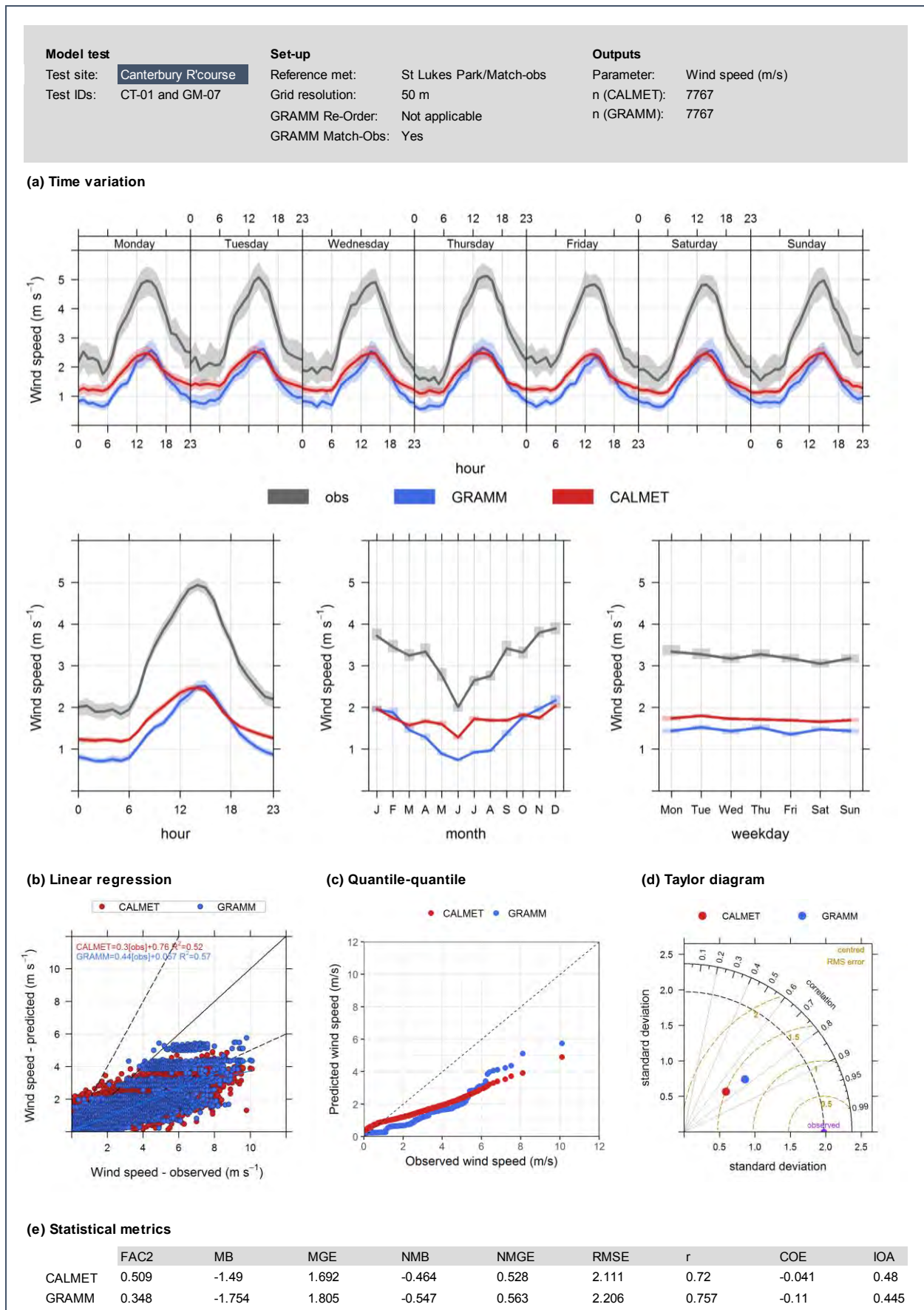


Figure E-32: Meteorological model performance – Canterbury Racecourse extract (tests CT-01 and GM-07)

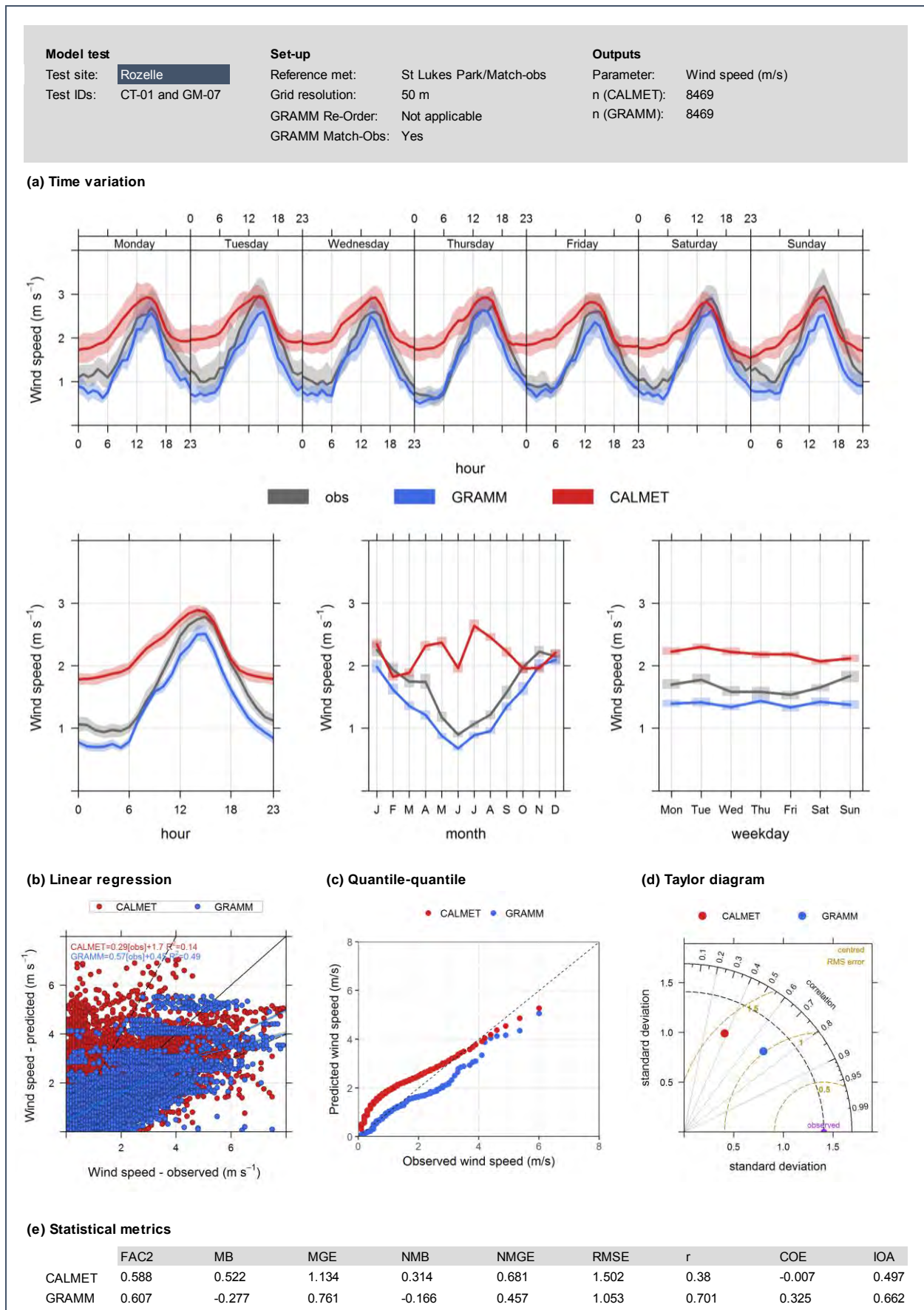


Figure E-33: Meteorological model performance – Rozelle extract (tests CT-01 and GM-07)

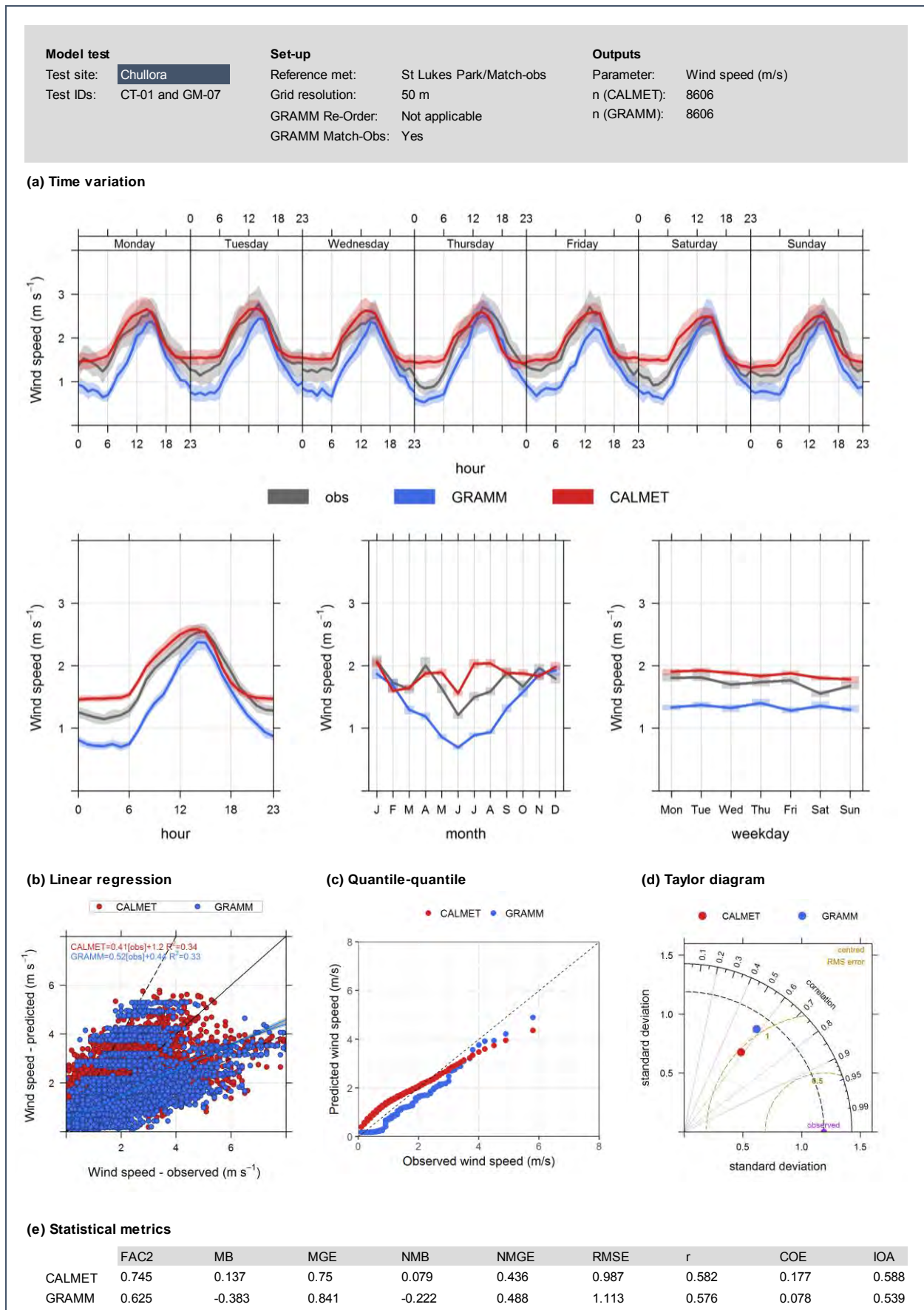


Figure E-34: Meteorological model performance – Chullora extract (tests CT-01 and GM-07)

E2 Wind direction evaluation

Figures E-35 to E-39 show the wind direction distributions as radar plots for all monitoring stations and tests. The plots simplify the presentation of the wind direction results, and allow observations and model performance to be compared more directly for each test. The results are summarised for St Lukes Park and the other stations separately below.

- St Lukes Park
 - In test CT-01 the CALMET extract was almost identical to the observed meteorology. As with the wind speed tests, this is not surprising as the station was used as input into CALMET.
 - The initial GRAMM test (GM-01) – with a grid spacing of 50 metres - gave a good representation of the wind direction distribution. However, the increased grid spacing of 100 metres in test GM-02 resulted in weaker GRAMM performance, with a marked underestimation of the frequency of northerly winds. Perhaps surprisingly, the results for test GM-03 (grid spacing of 200 metres) were quite similar to those from test GM-01.
 - GRAMM test GM-04 (with Re-Order) gave very good results.
 - The performance of CALMET in test CT-02 (no reference meteorology for St Lukes Park) was markedly worse than in test CT-01. However, tests GM-05 (Match-to-Observations at all stations except St Lukes Park) and GM-06 (Match-to-Observations at all stations) also performed quite poorly.
 - Test GM-07 (Match-to-Observations at St Lukes Park only) gave good results.
- Other stations
 - In test CT-01 the reference meteorology was taken from St Lukes Park only, and the performance of CALMET at all the other stations was clearly poorer than at St Lukes Park.
 - For tests GM-01 to GM-04, the results from GRAMM were generally quite poor at all stations, and there was often a considerable overestimation of northerly winds.
 - Test CT-02 resulted in near-perfect CALMET performance, again unsurprisingly.
 - In tests GM-05 and GM-06 GRAMM gave a fair to good performance.
 - Test GM-07 resulted in relatively poor GRAMM performance.

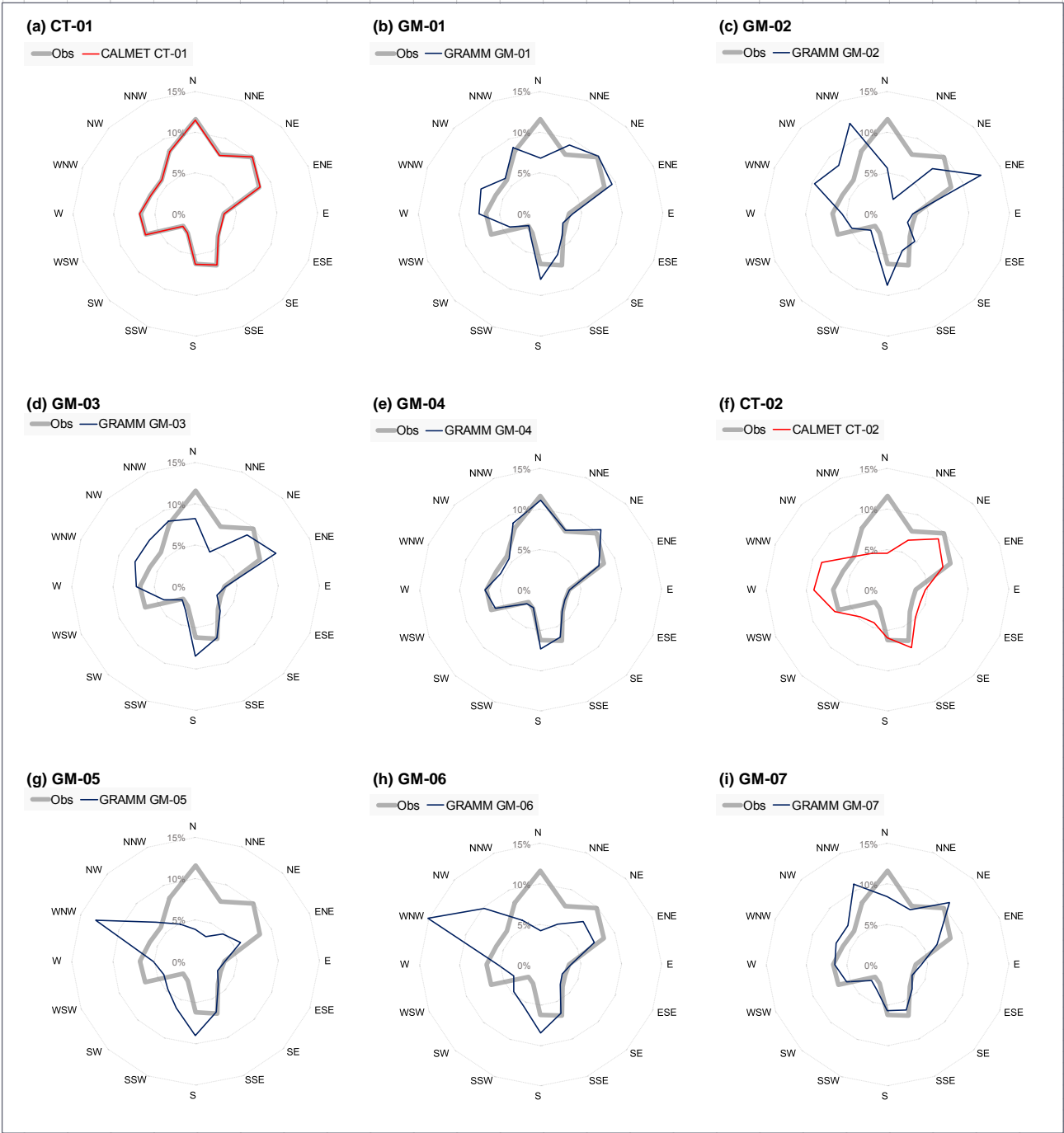


Figure E-35: Radar plots showing wind direction distribution by test – St Lukes Park extract

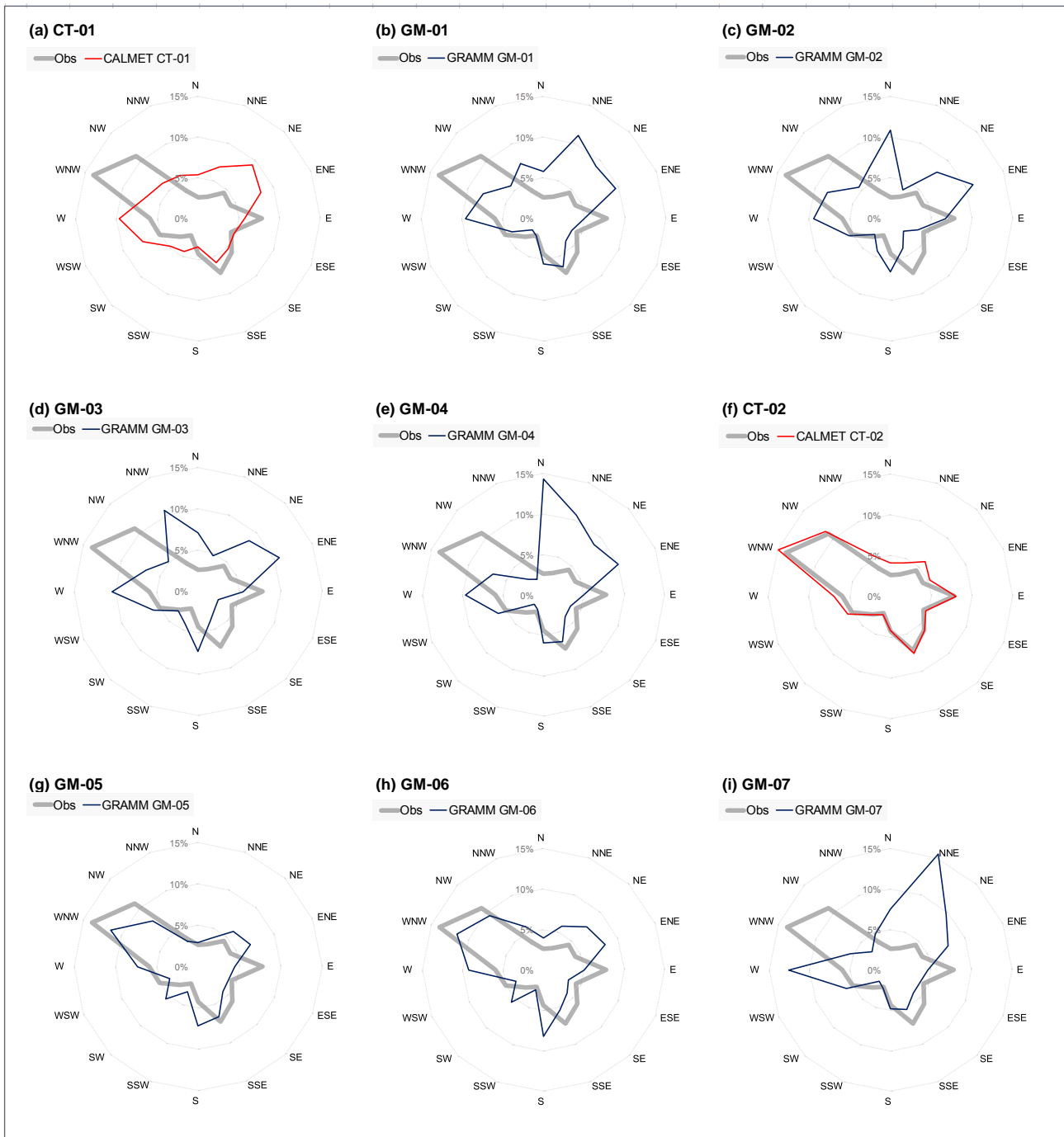


Figure E-36: Radar plots showing wind direction distribution by test – Olympic Park extract

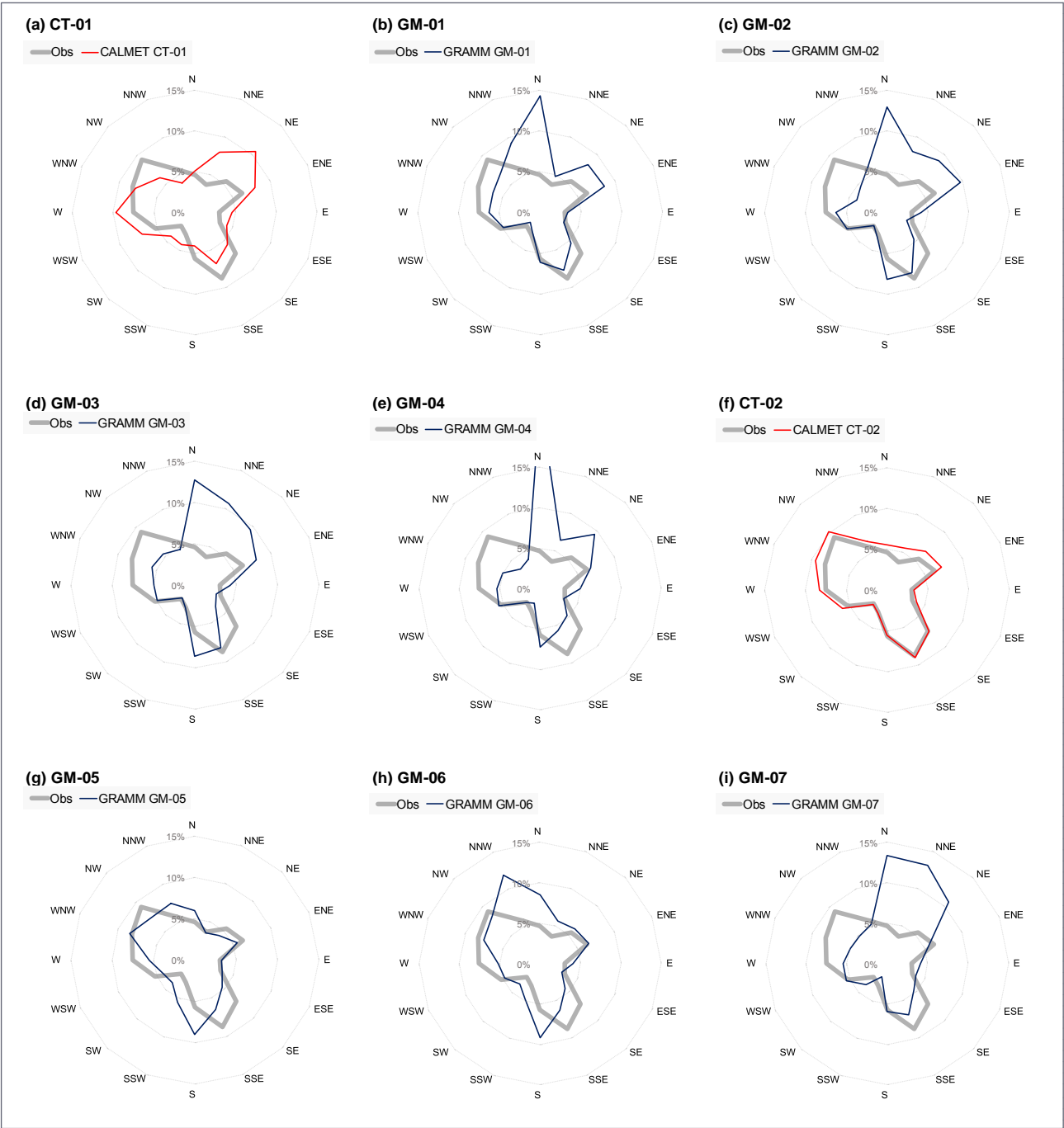


Figure E-37: Radar plots showing wind direction distribution by test – Canterbury Racecourse extract

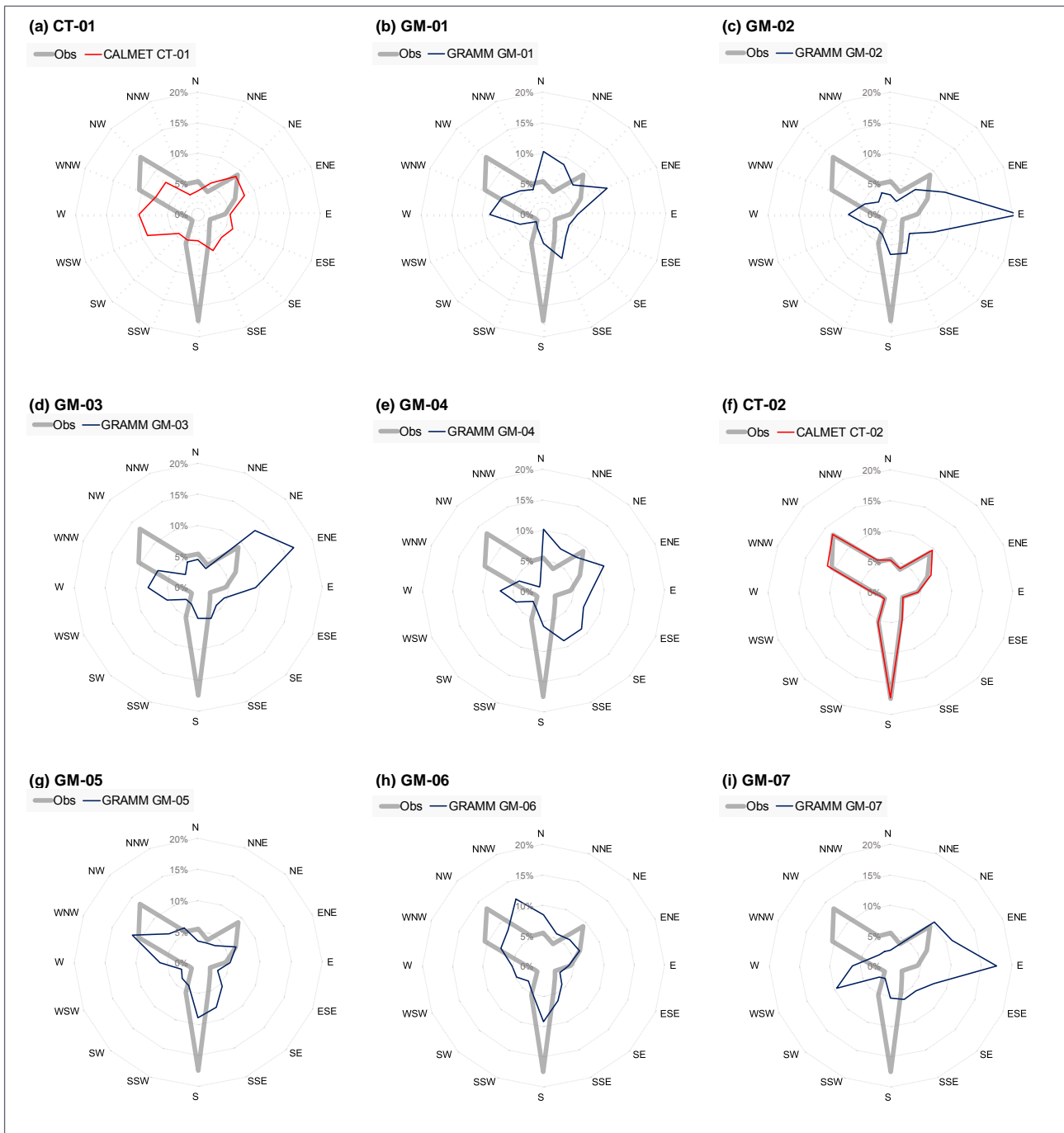


Figure E-38: Radar plots showing wind direction distribution by test – Rozelle extract

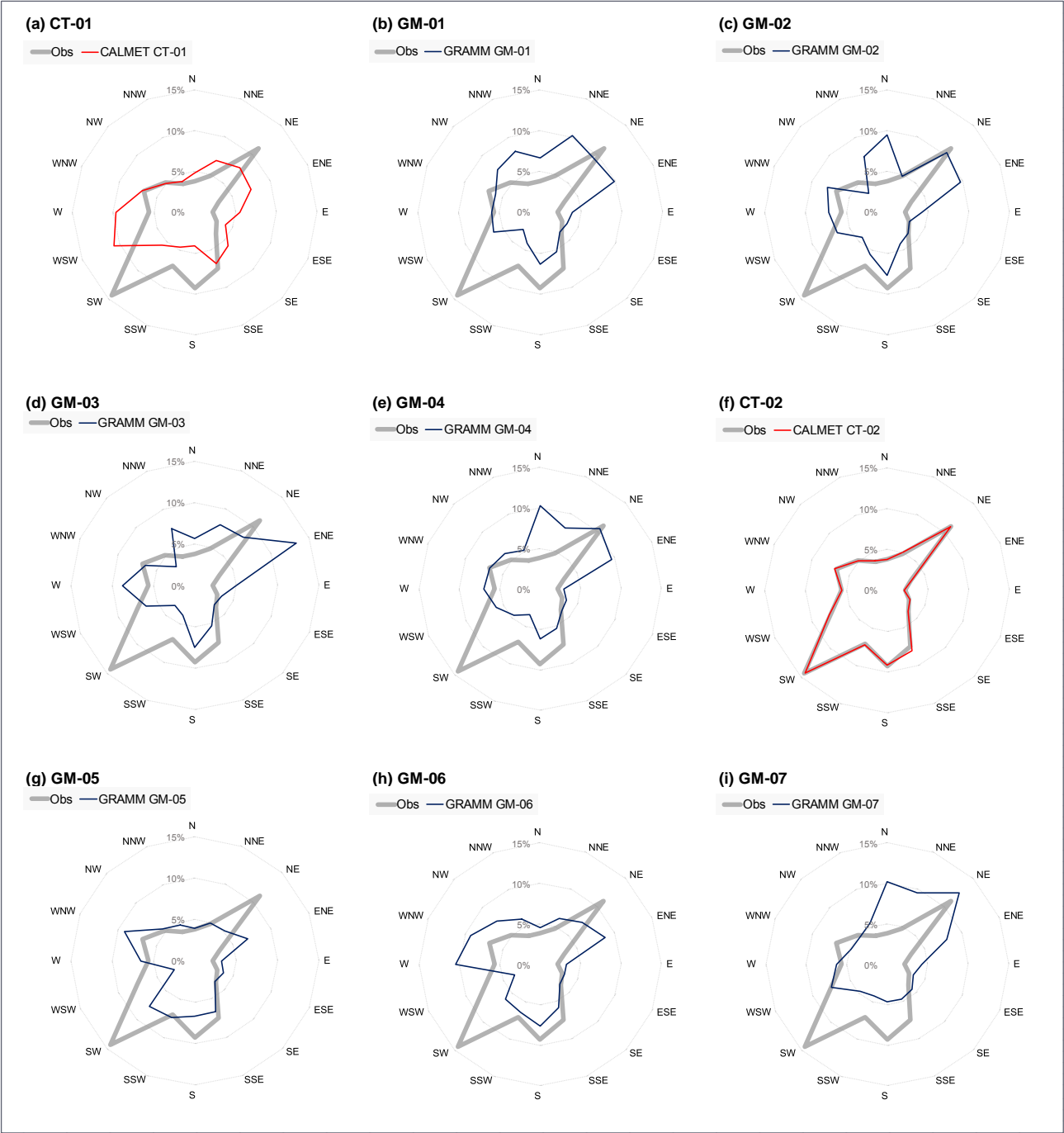
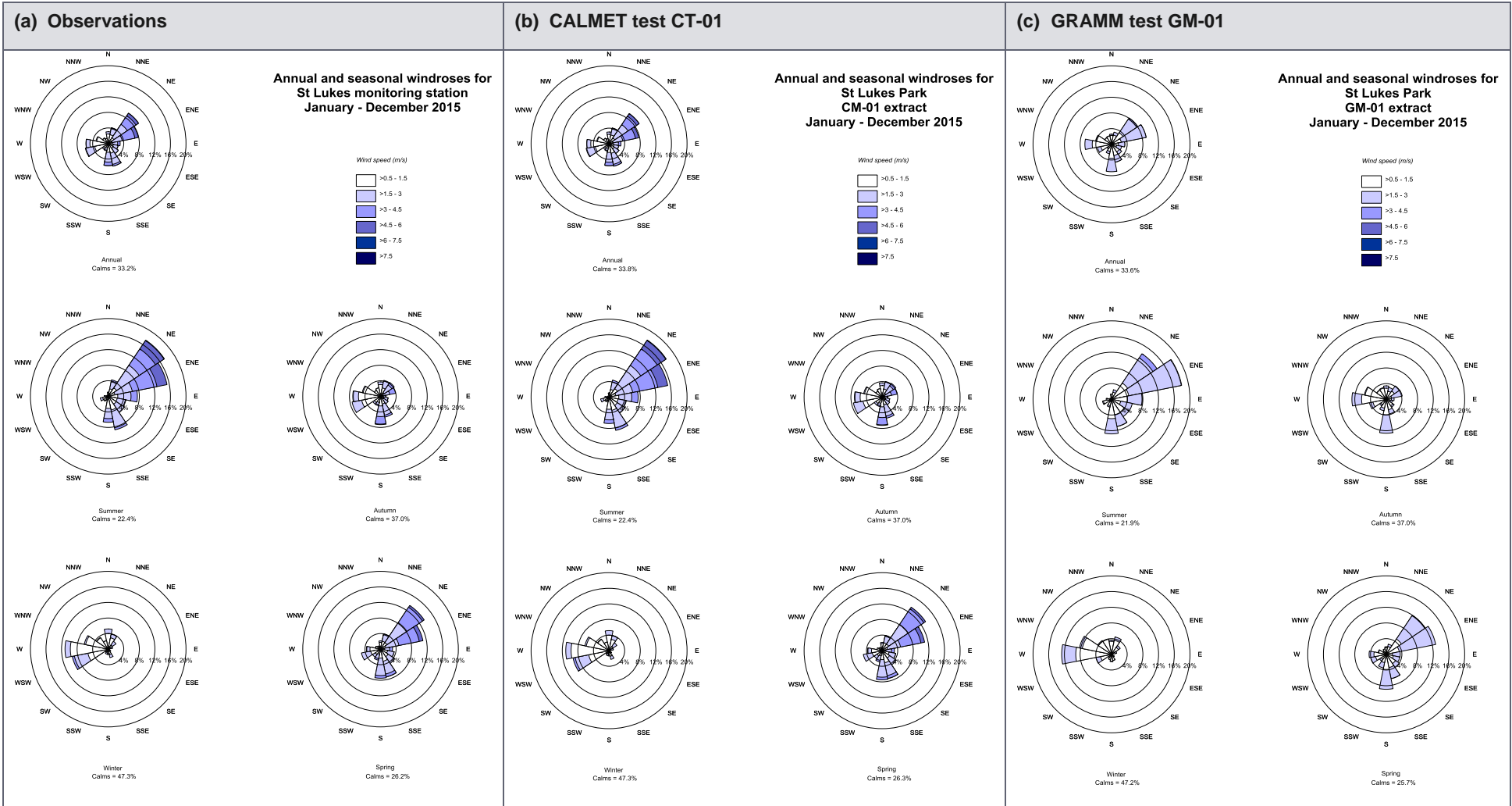


Figure E-39: Radar plots showing wind direction distribution by test – Chullora extract

E3 Wind roses

On the following pages annual and seasonal wind roses are shown for all tests. These have not been analysed in detail, and are provided for reference.



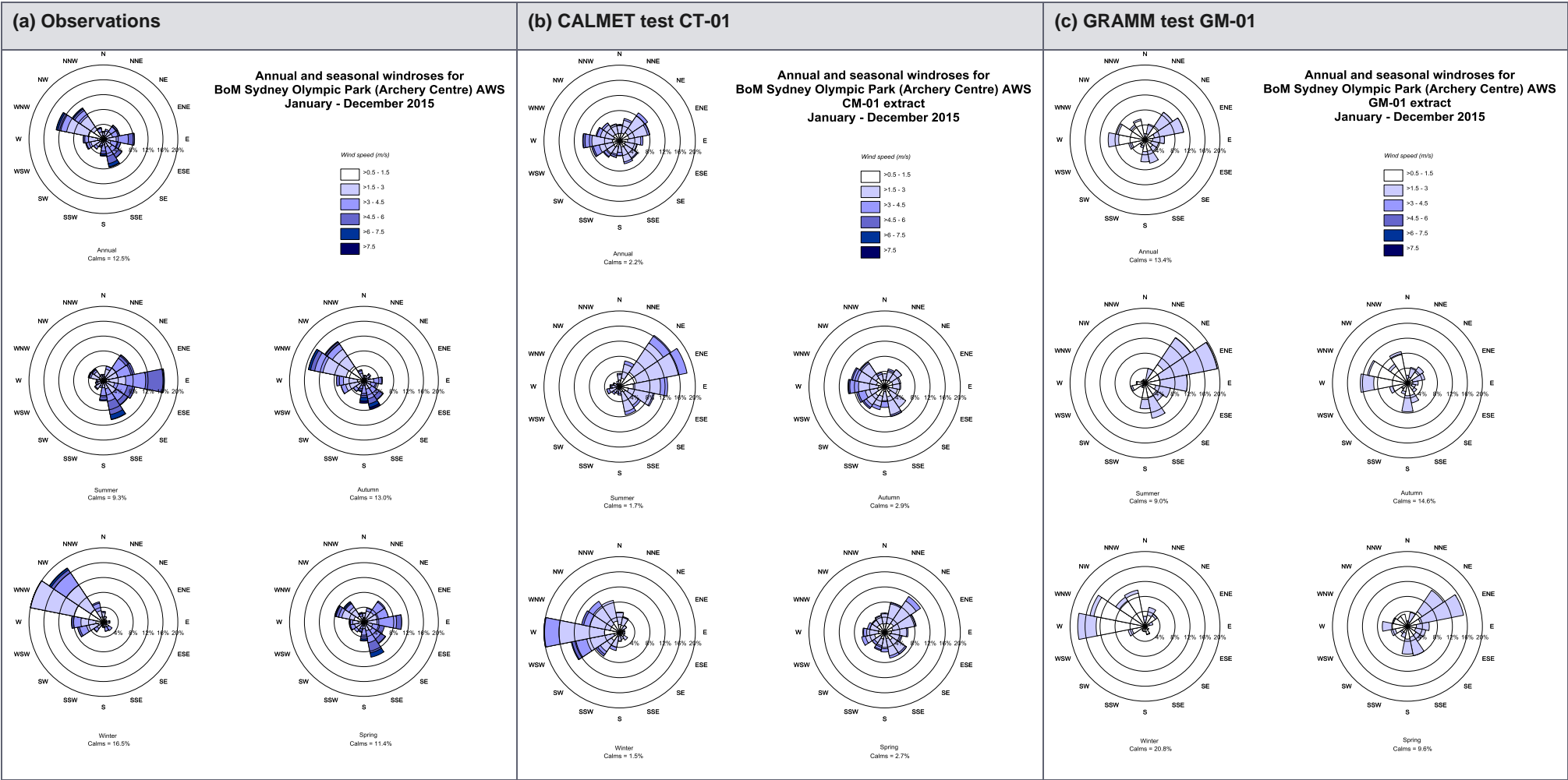


Figure E-41: Annual and seasonal wind roses - Olympic Park (observations, test CT-01 and test GM-01)

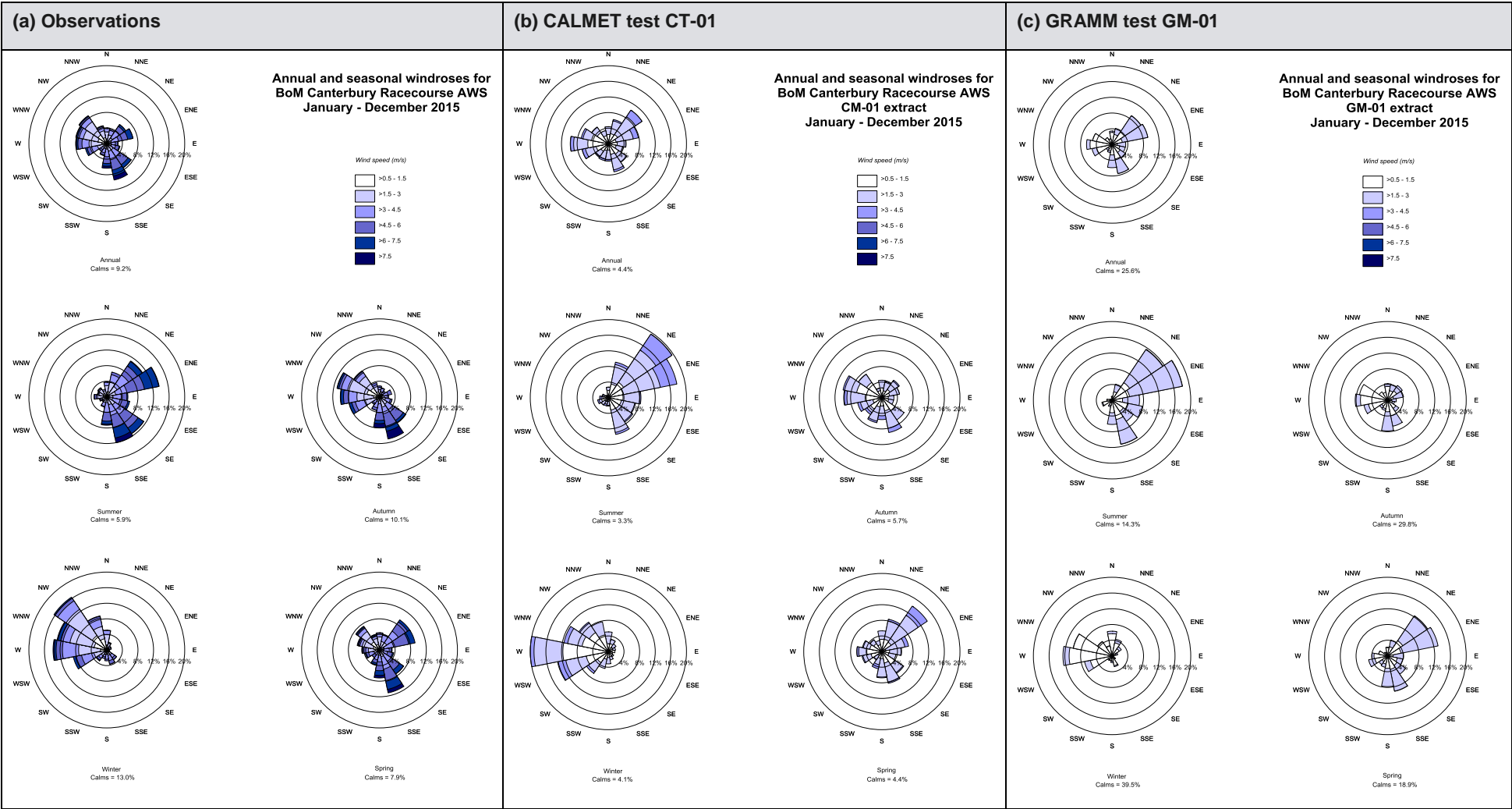
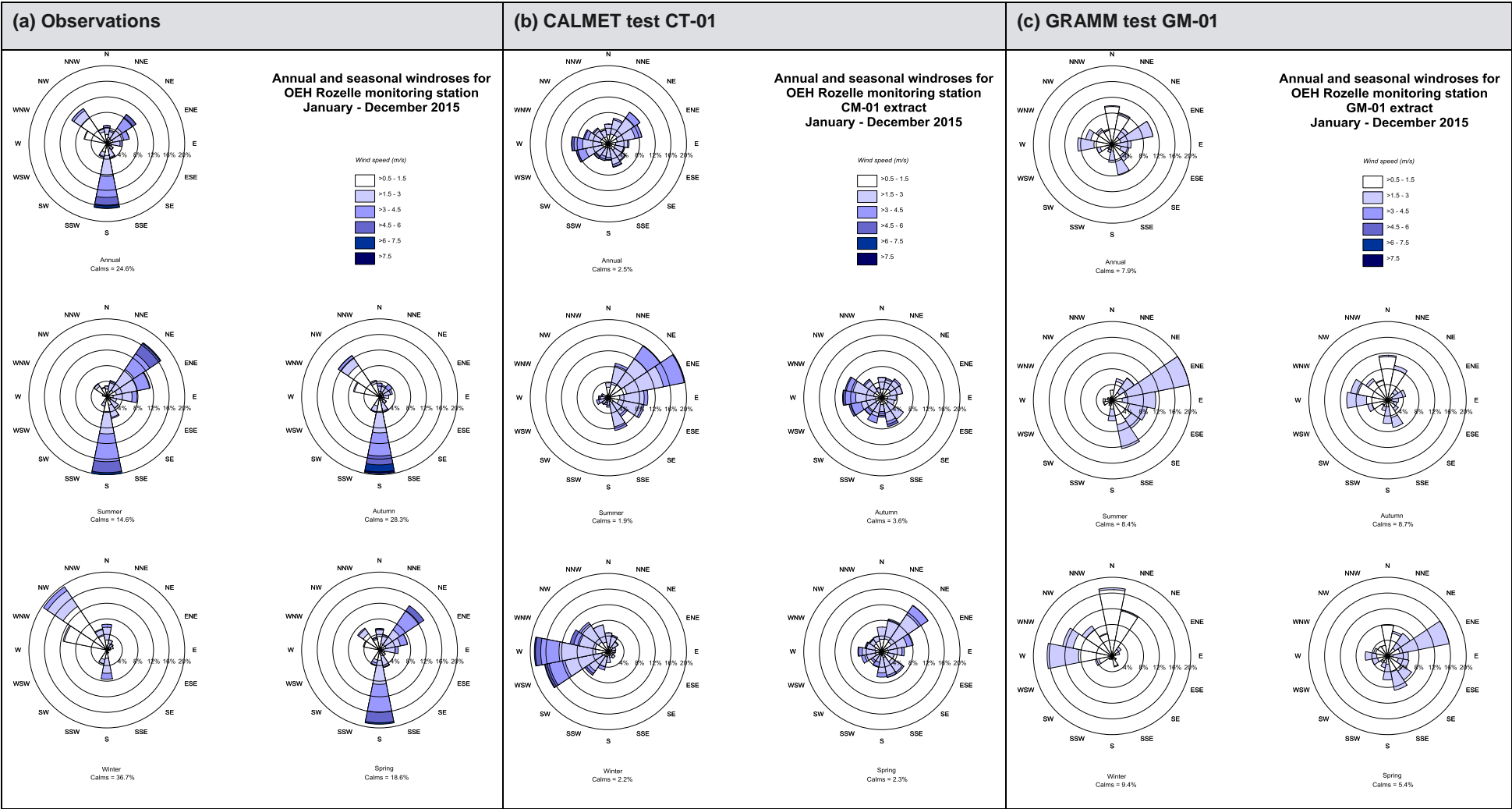


Figure E-42: Annual and seasonal wind roses - Canterbury Racecourse (observations, test CT-01 and test GM-01)



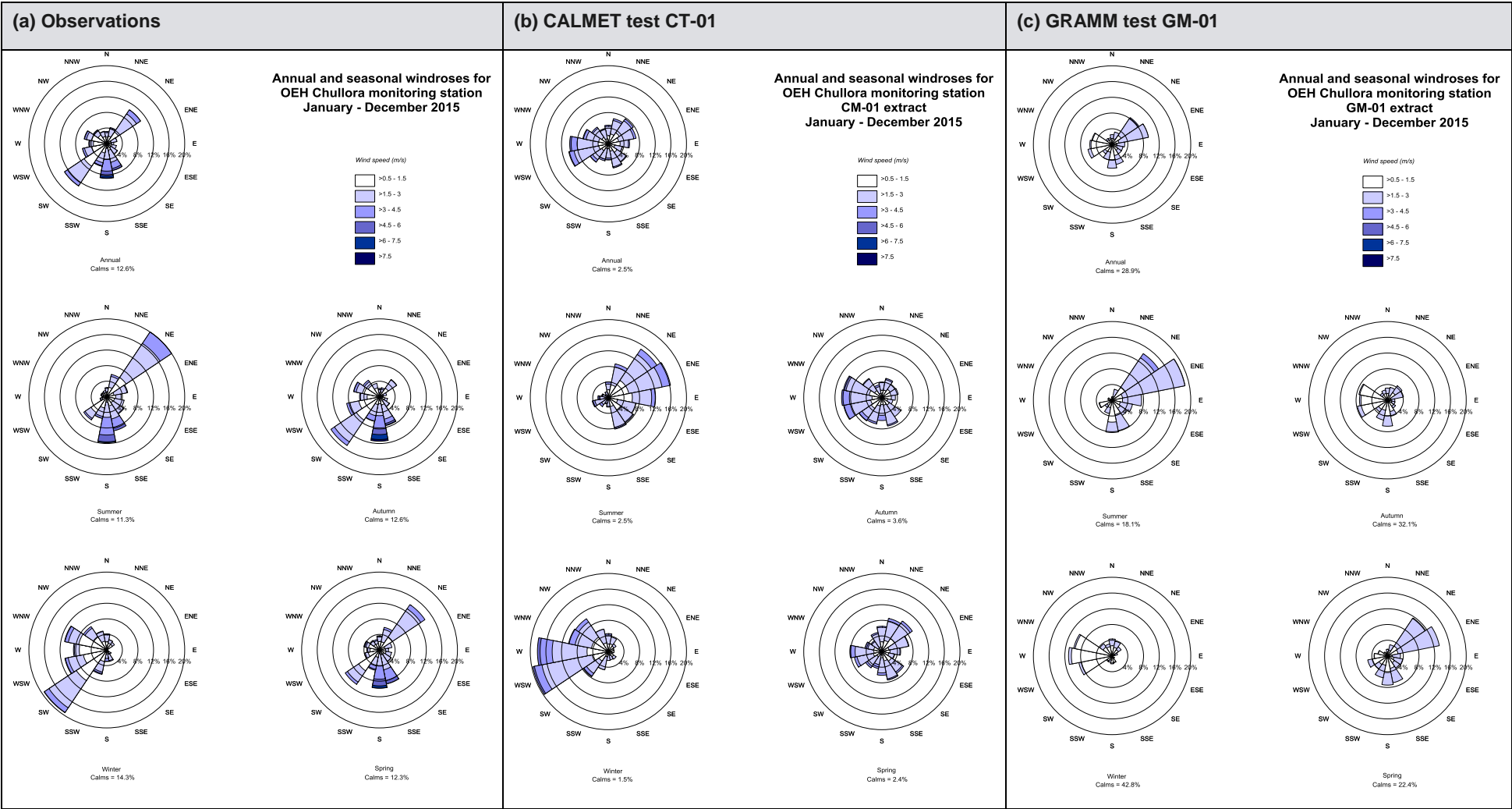
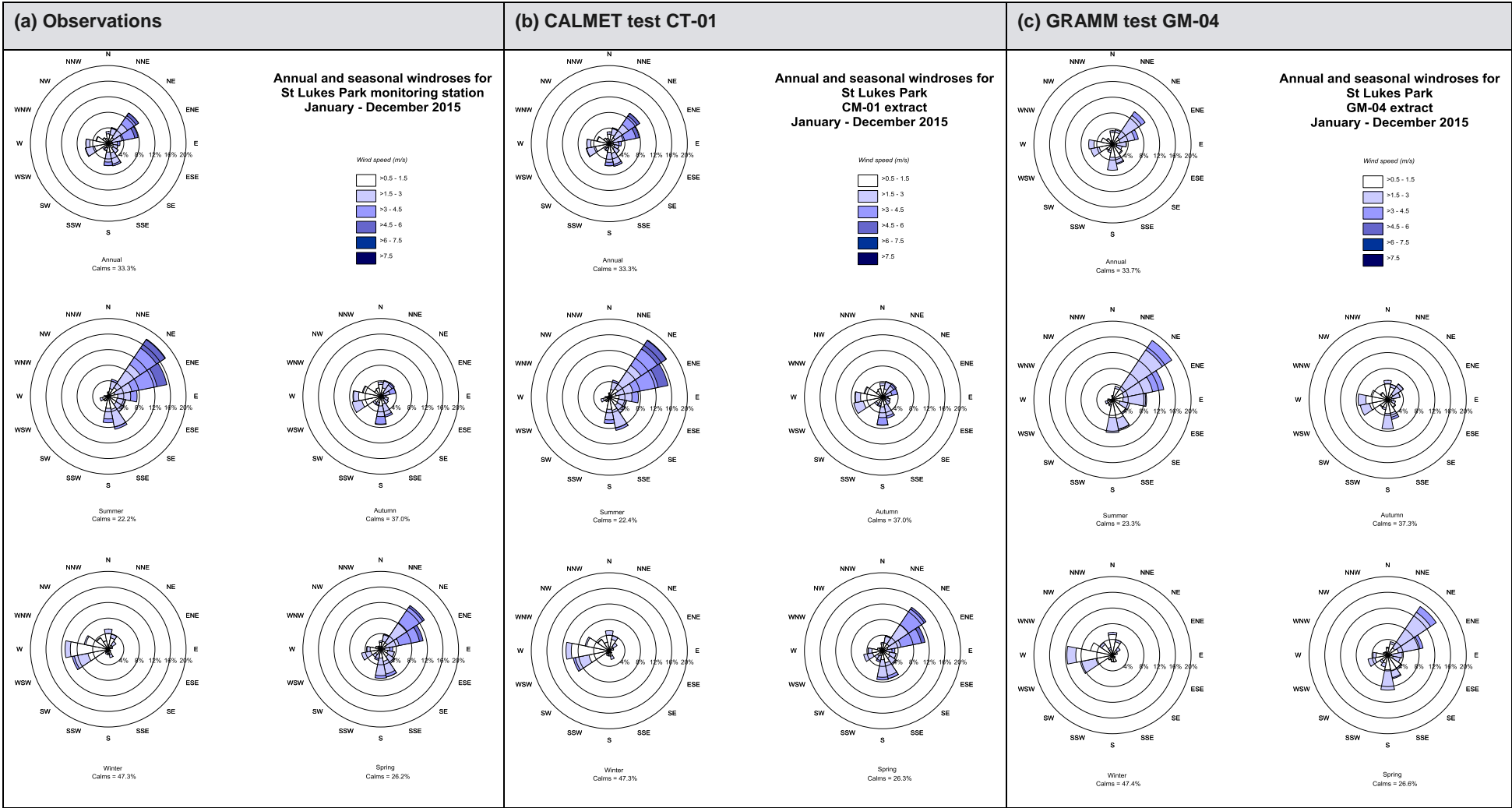


Figure E-44: Annual and seasonal wind roses – Chullora (observations, test CT-01 and test GM-01)



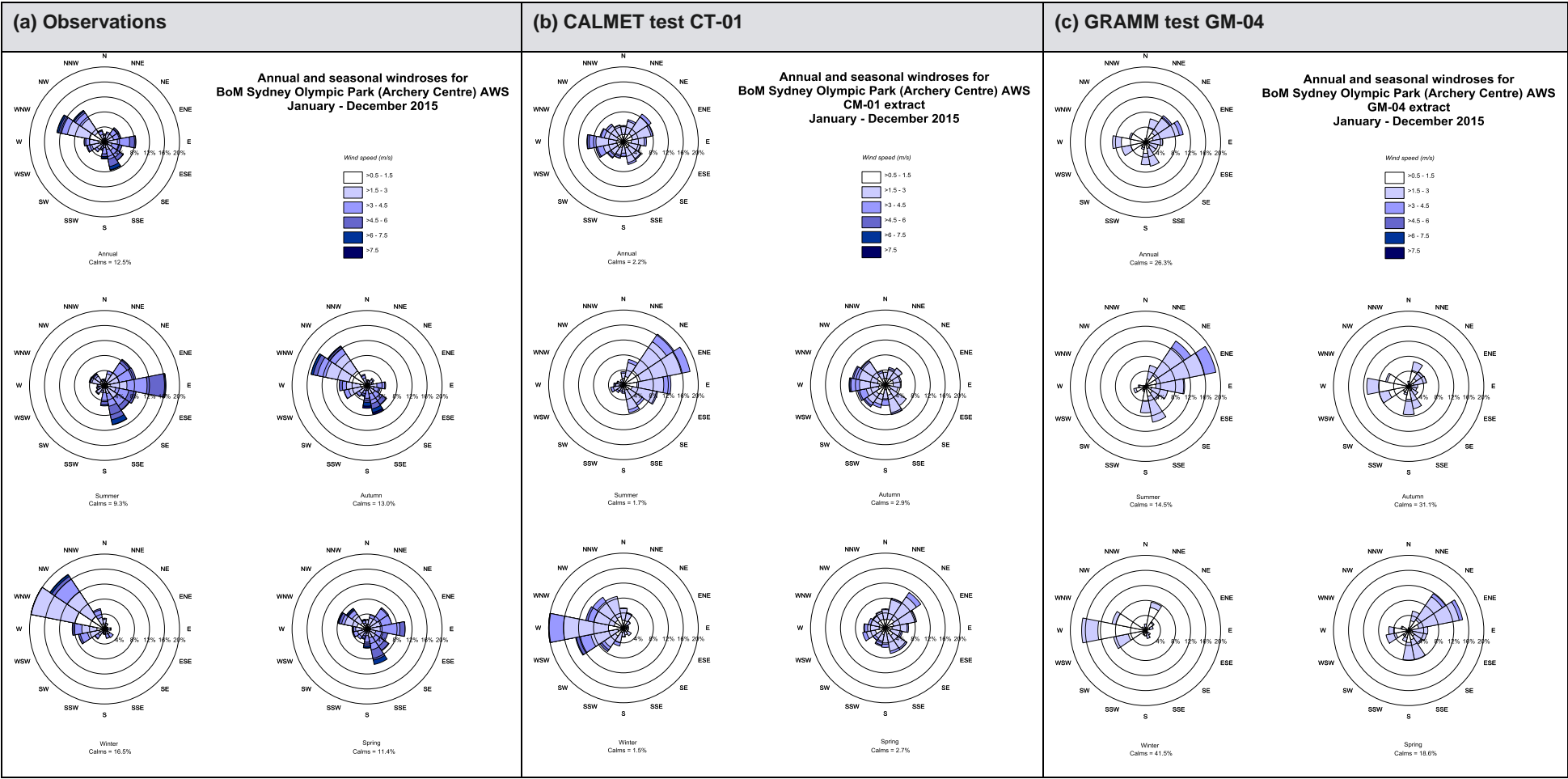


Figure E-46: Effects of Re-Order in GRAMM – Sydney Olympic Park extract (observations, test CT-01 and test GM-04)

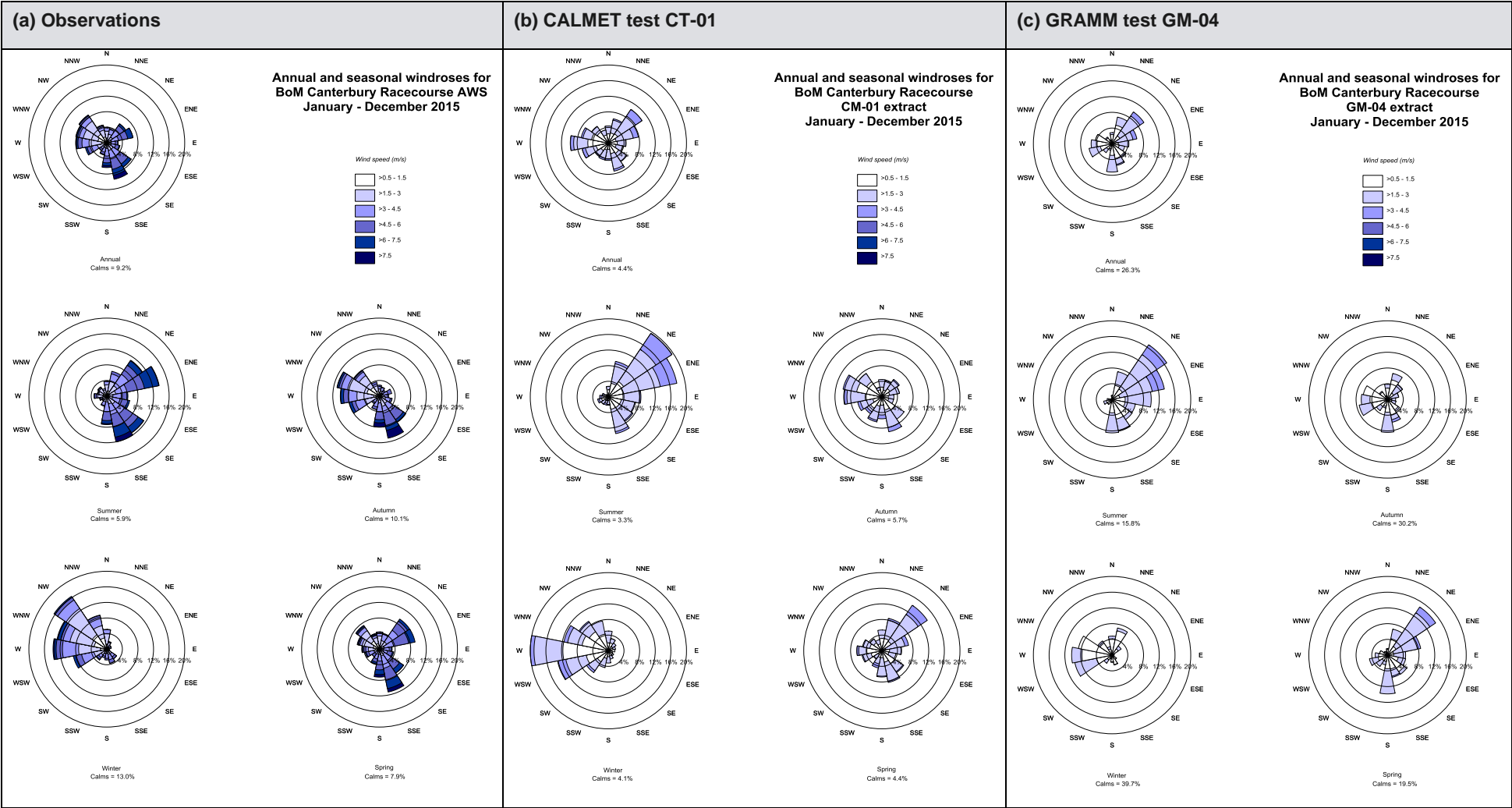
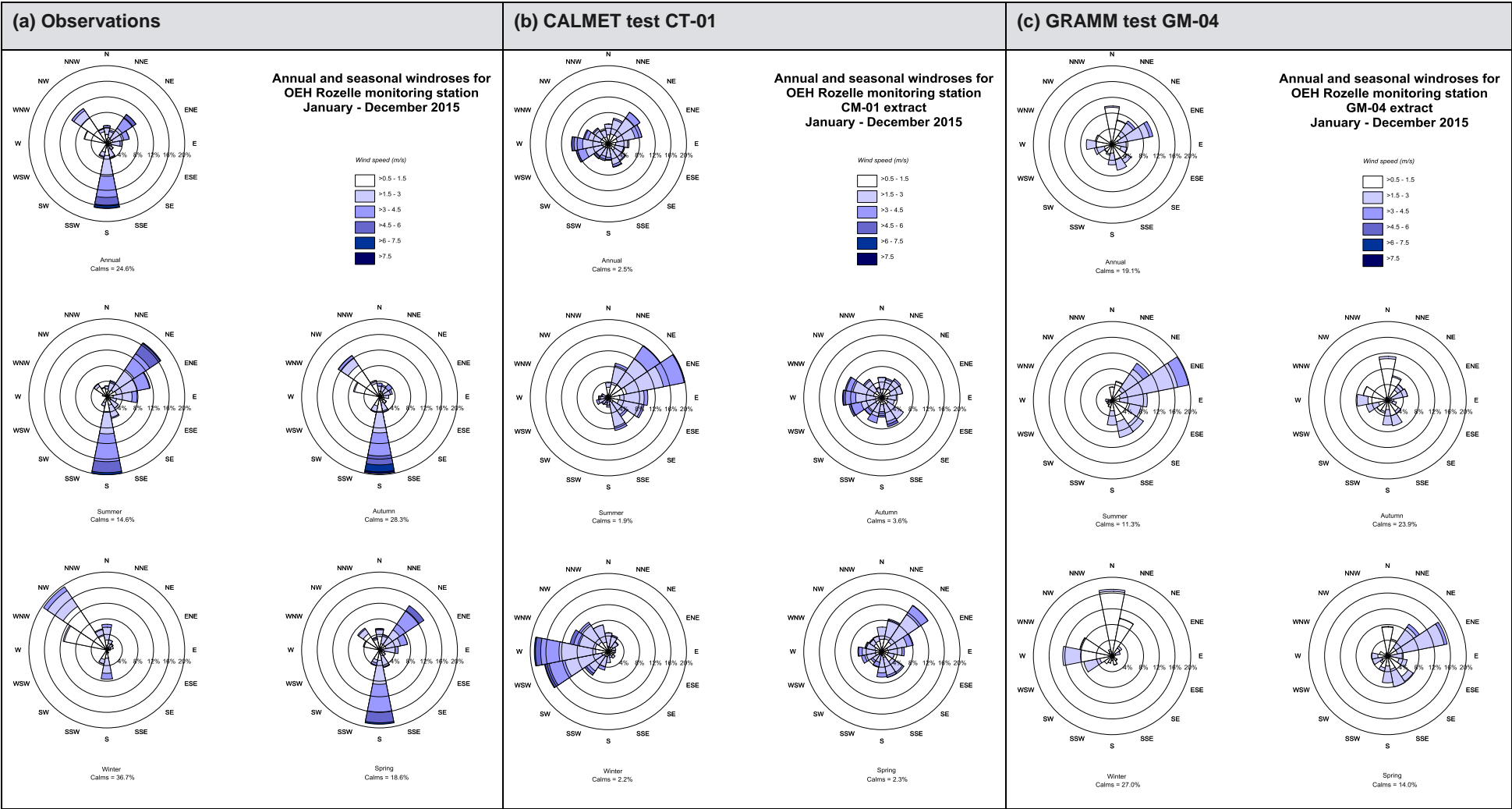


Figure E-47: Effects of Re-Order in GRAMM – Canterbury Racecourse extract (observations, test CT-01 and test GM-04)



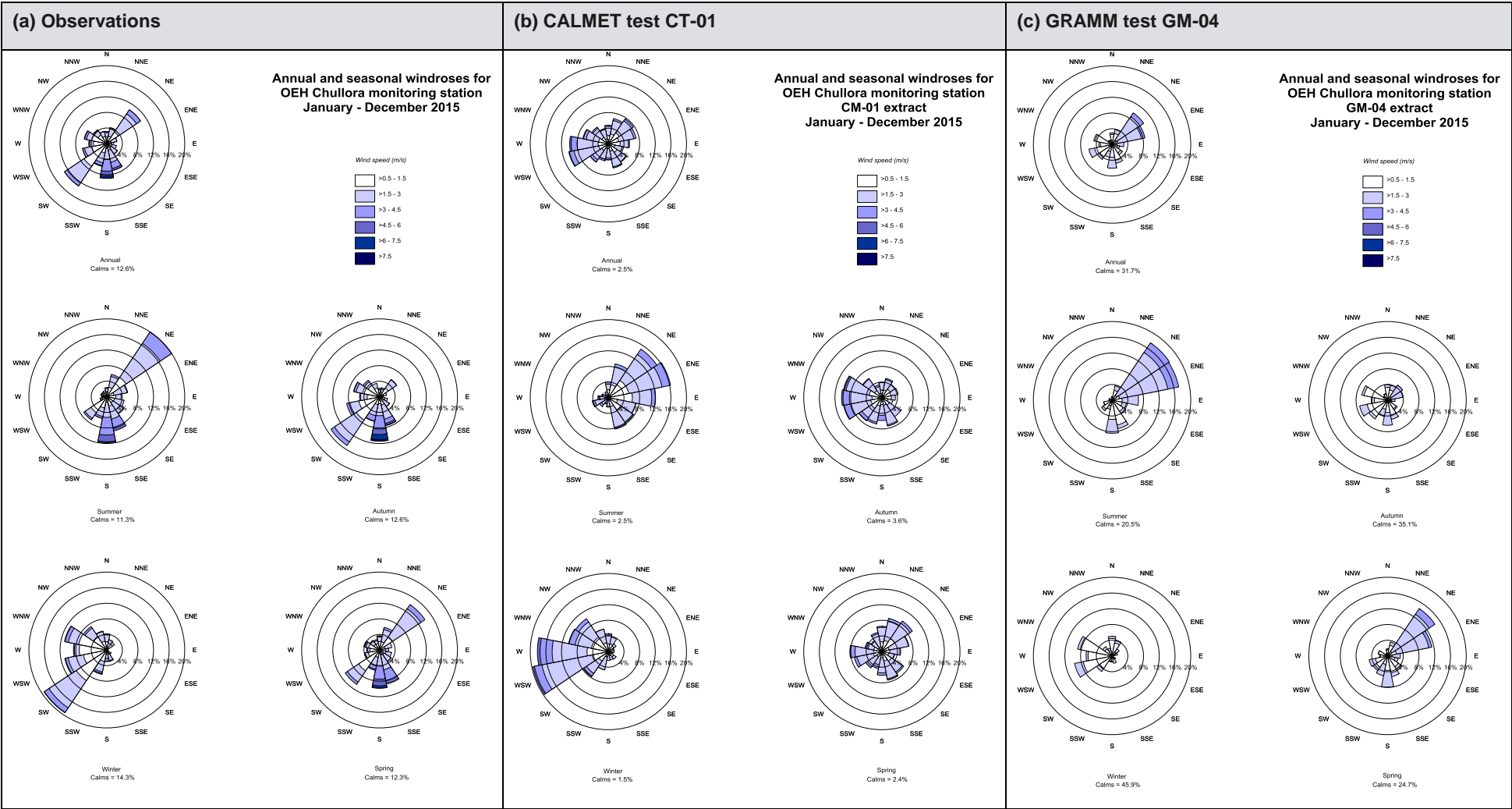


Figure E-49: Effects of Re-Order in GRAMM – Chullora extract (observations, test CT-01 and test GM-04)

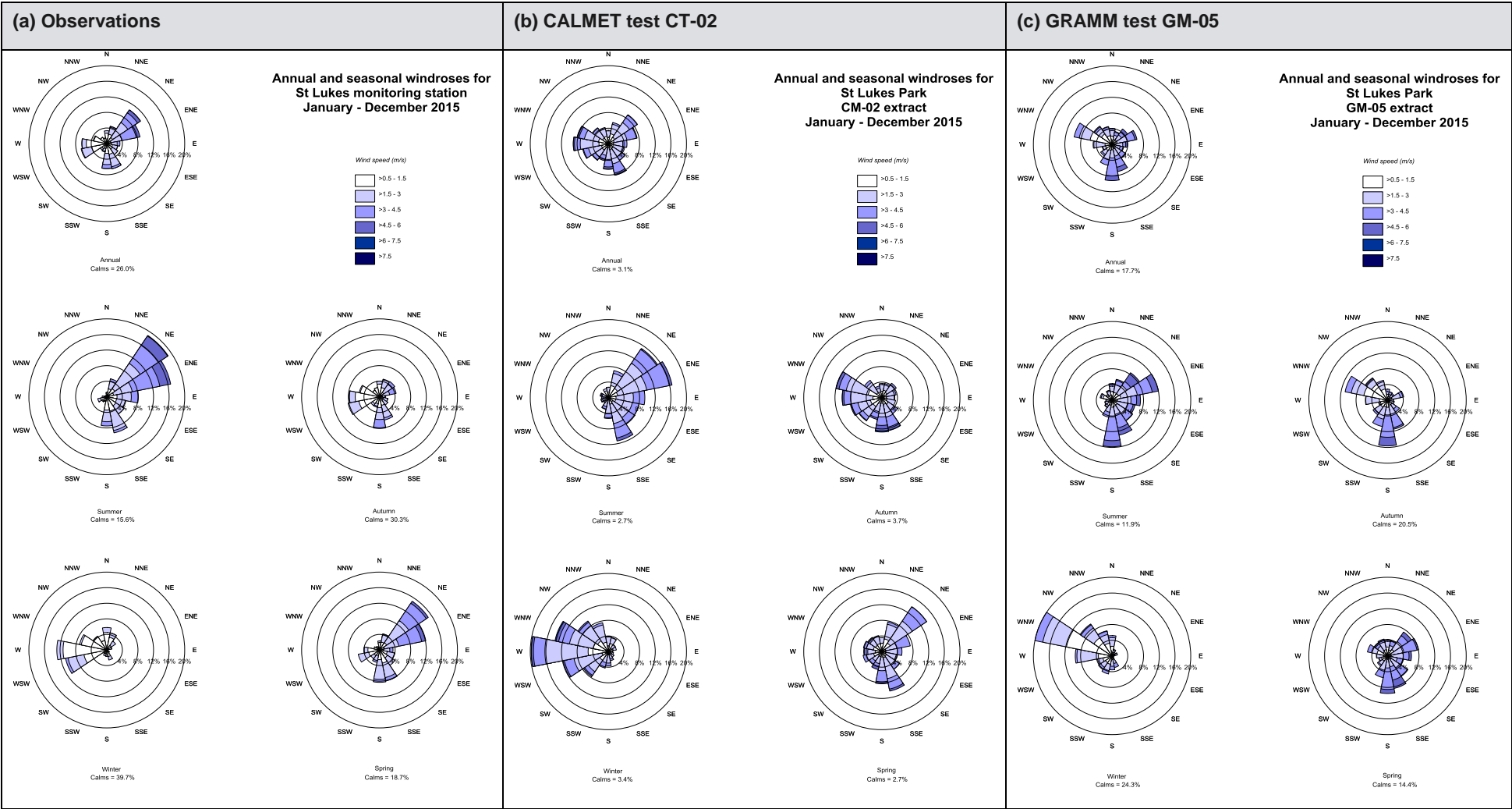


Figure E-50: Model performance - St Lukes Park extract (observations, test CT-02 and test GM-05)

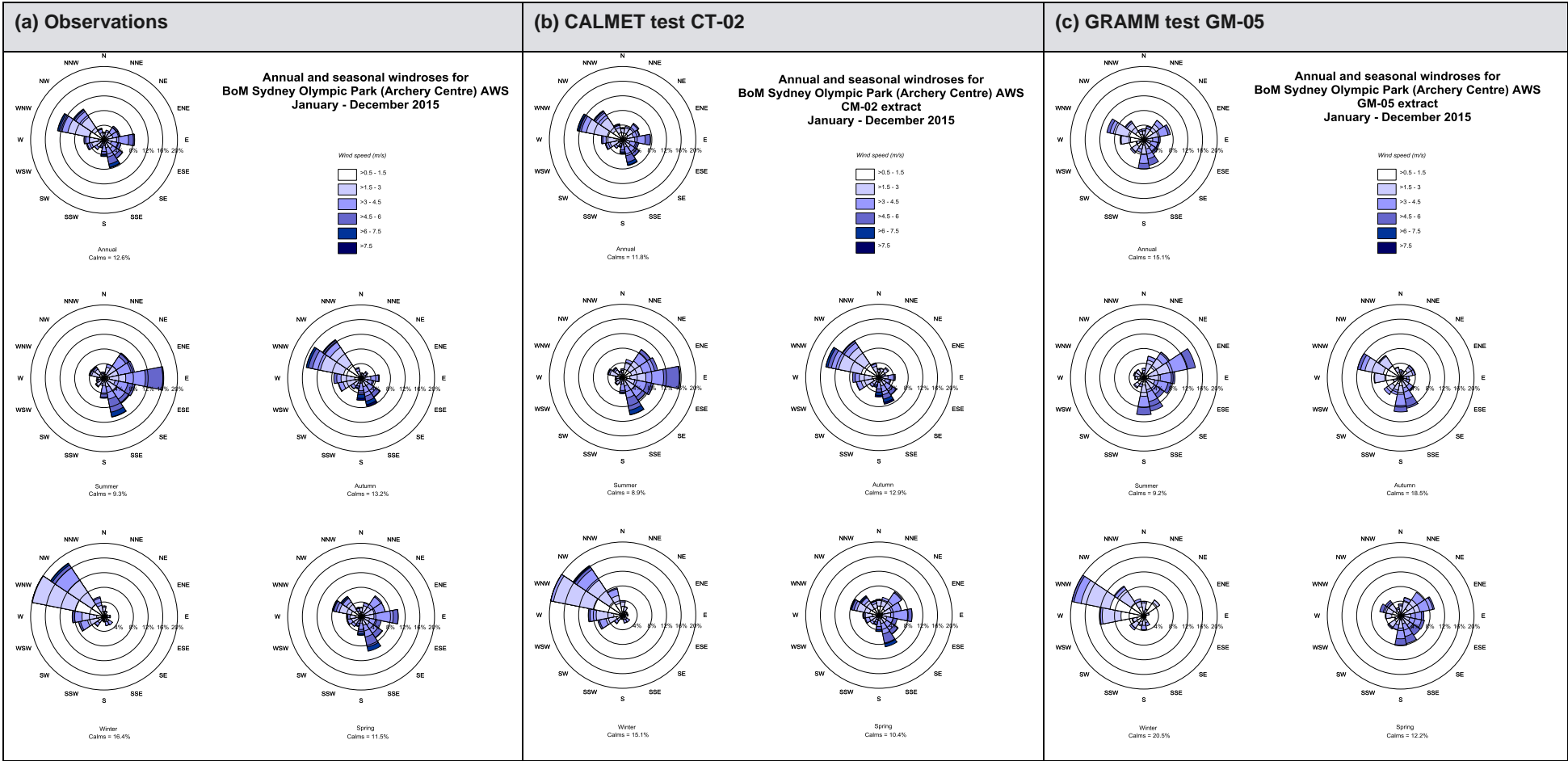


Figure E-51: Model performance – Sydney Olympic Park extract (observations, test CT-02 and test GM-05)

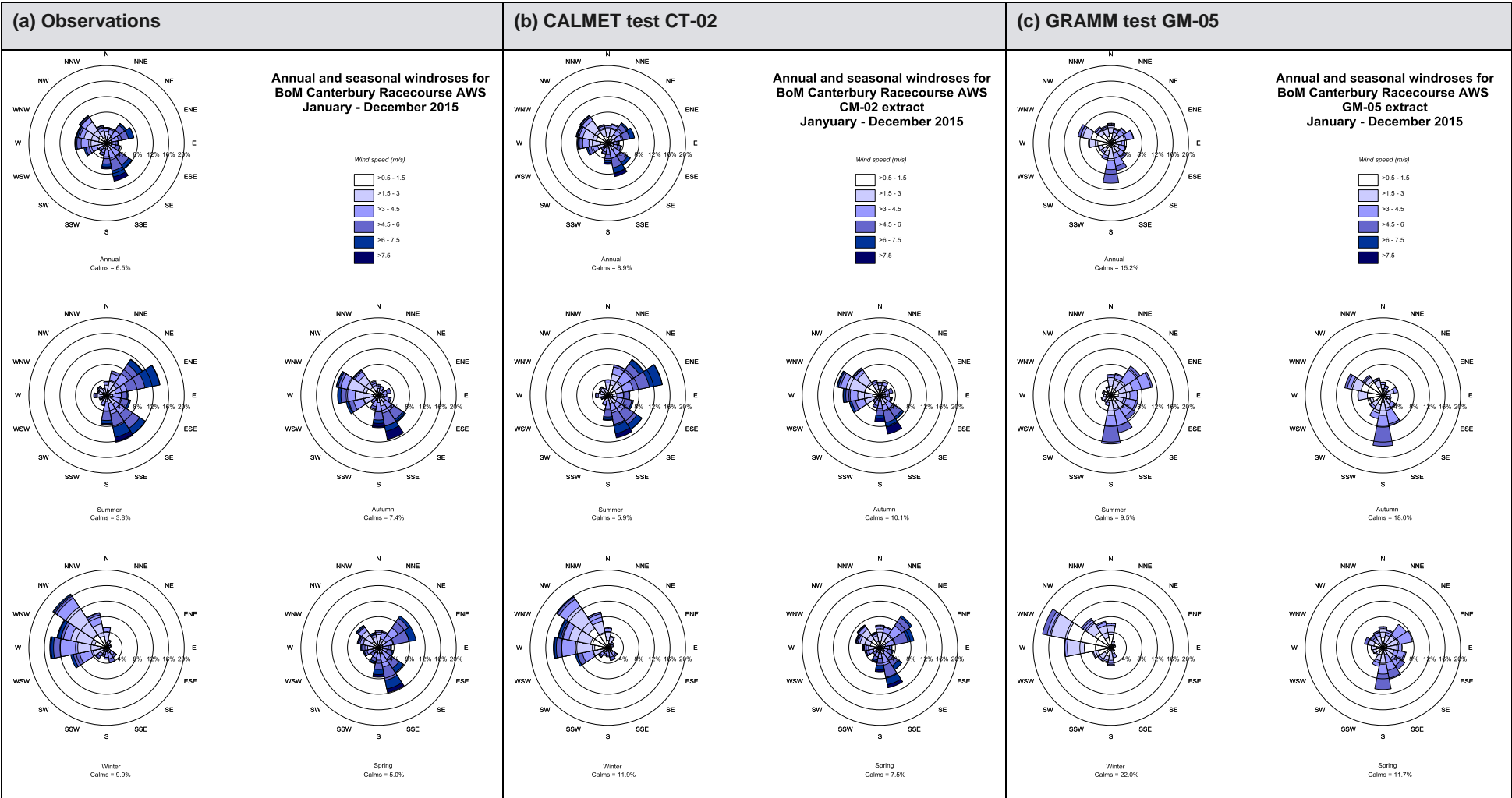
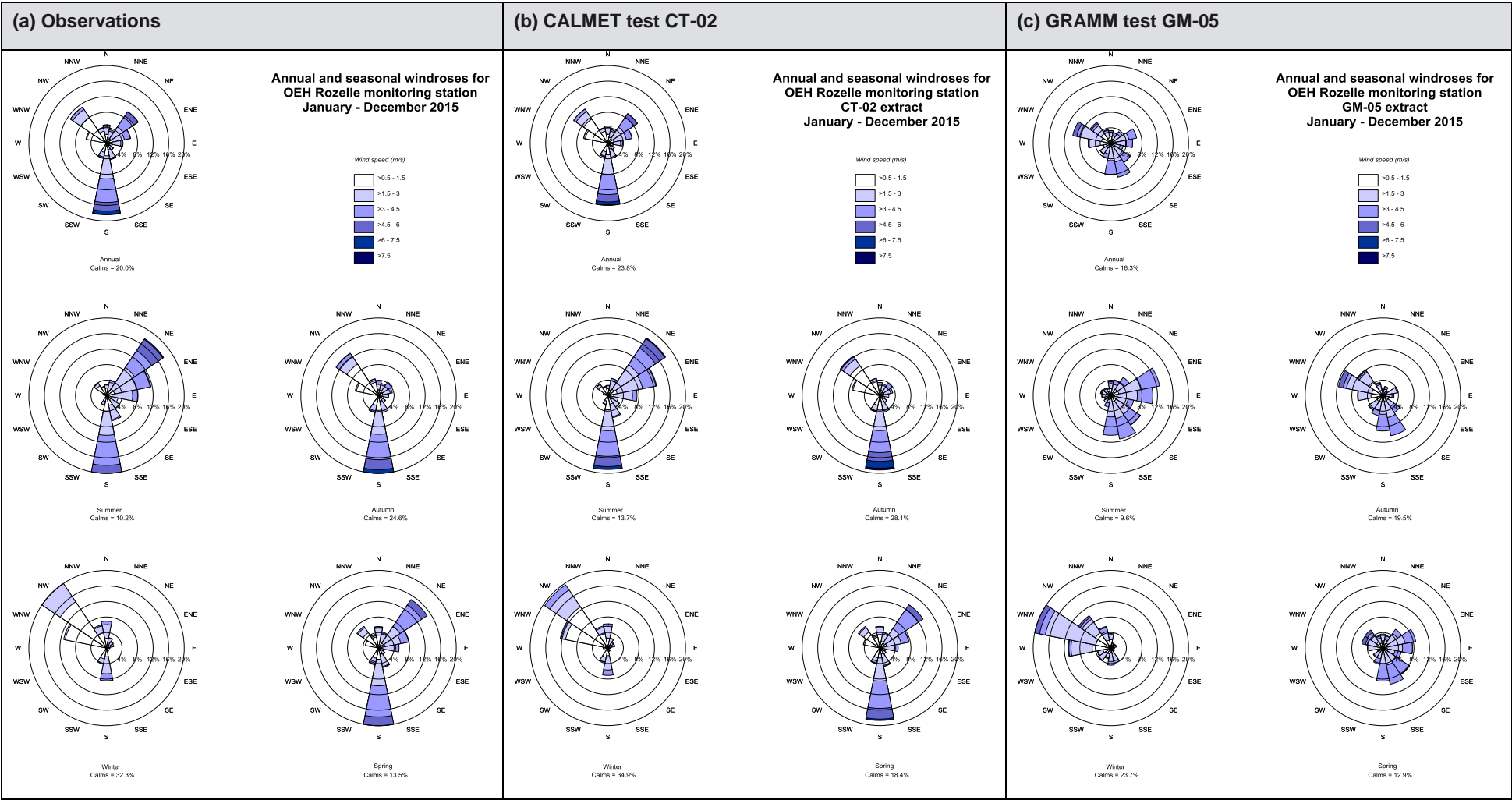
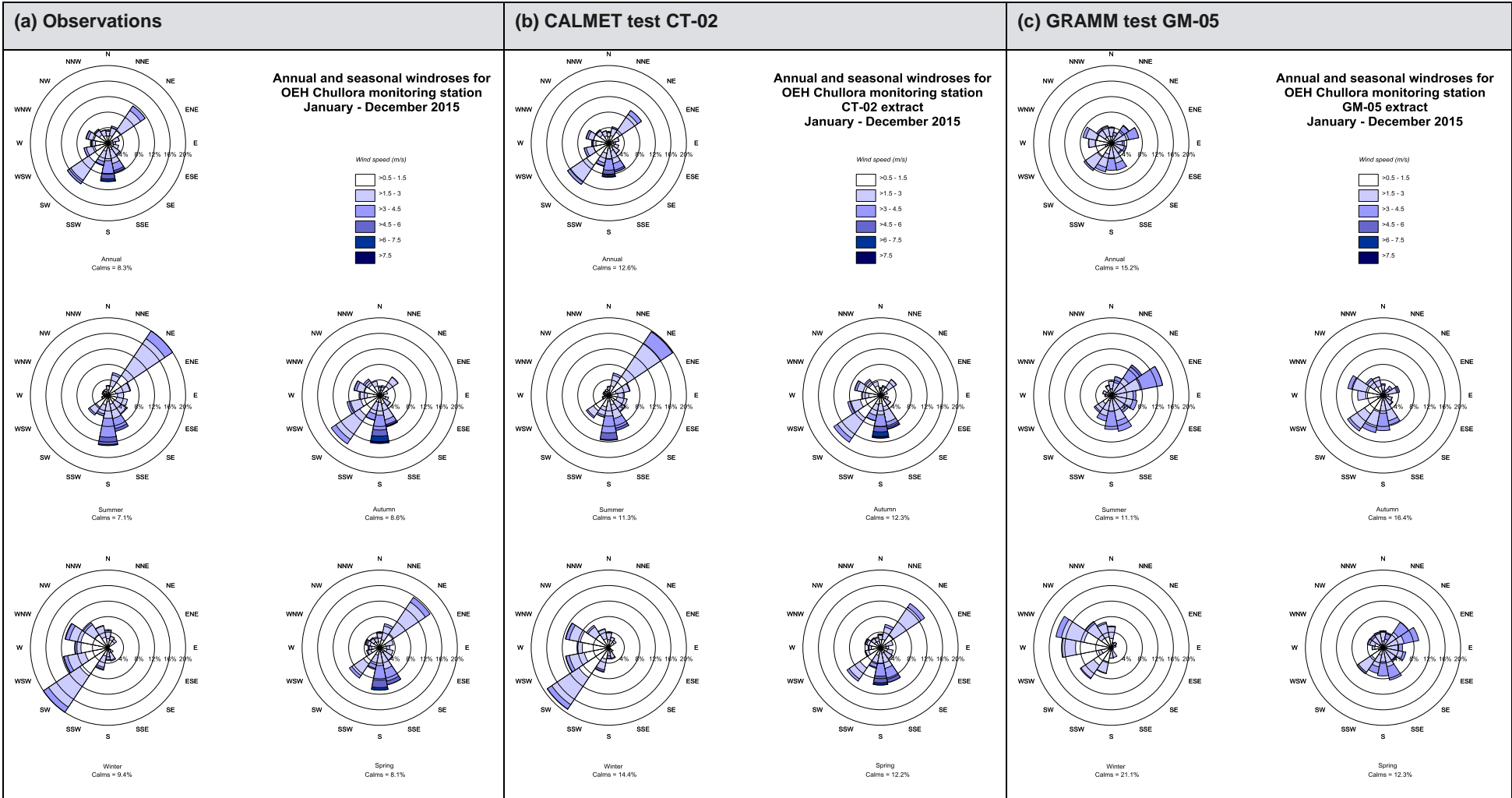


Figure E-52: Model performance – Canterbury Racecourse extract (observations, test CT-02 and test GM-05)





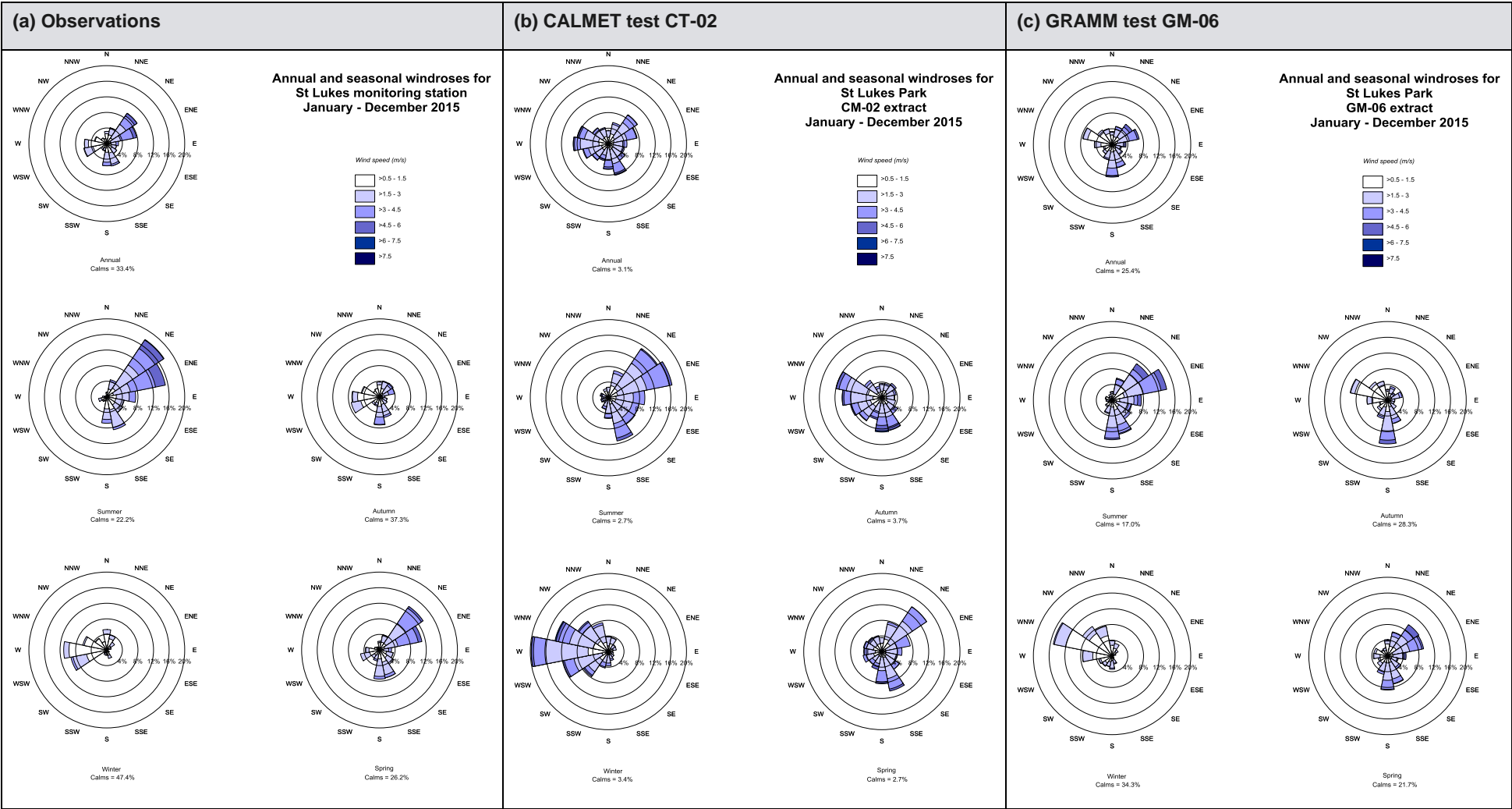
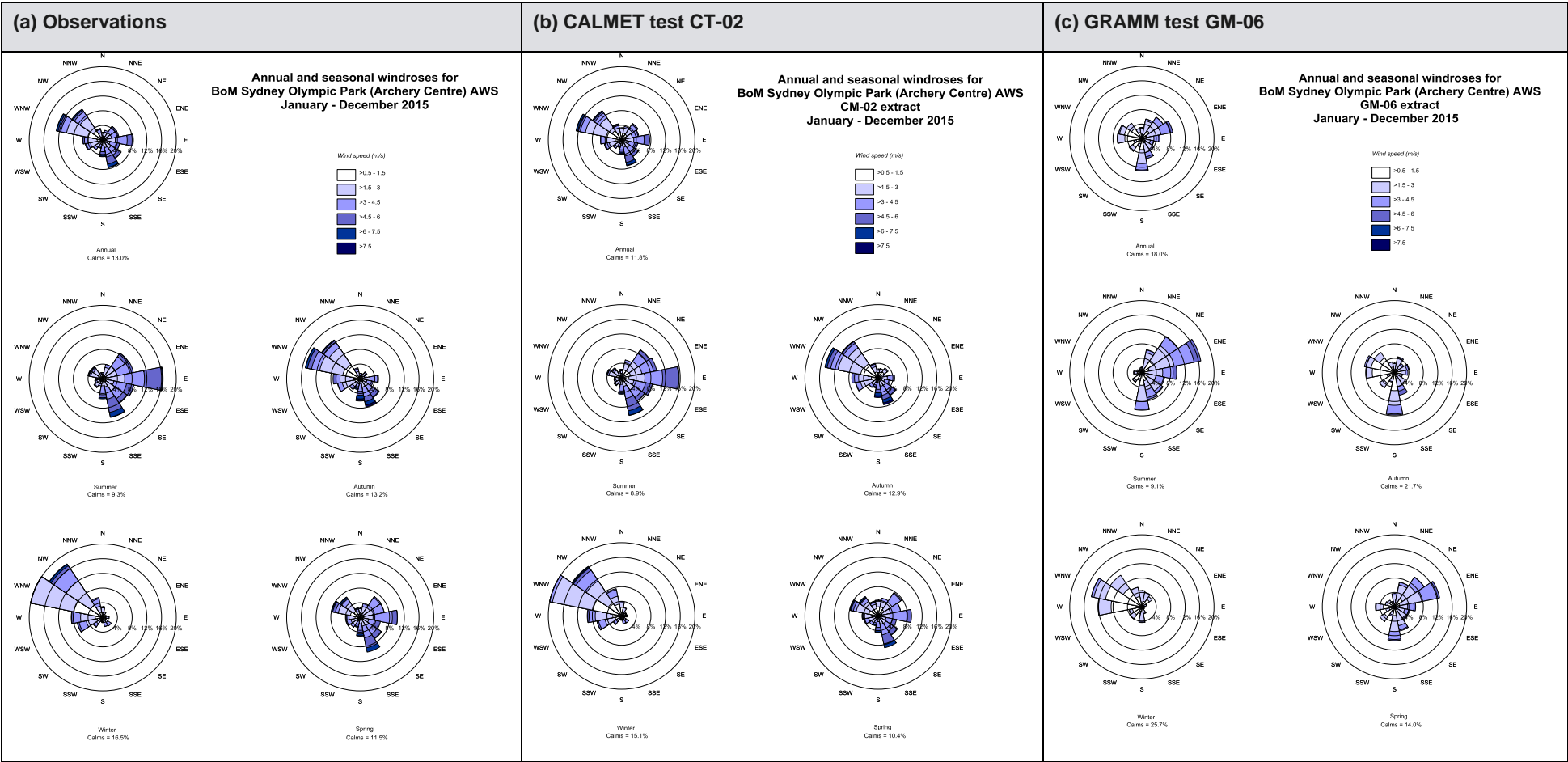
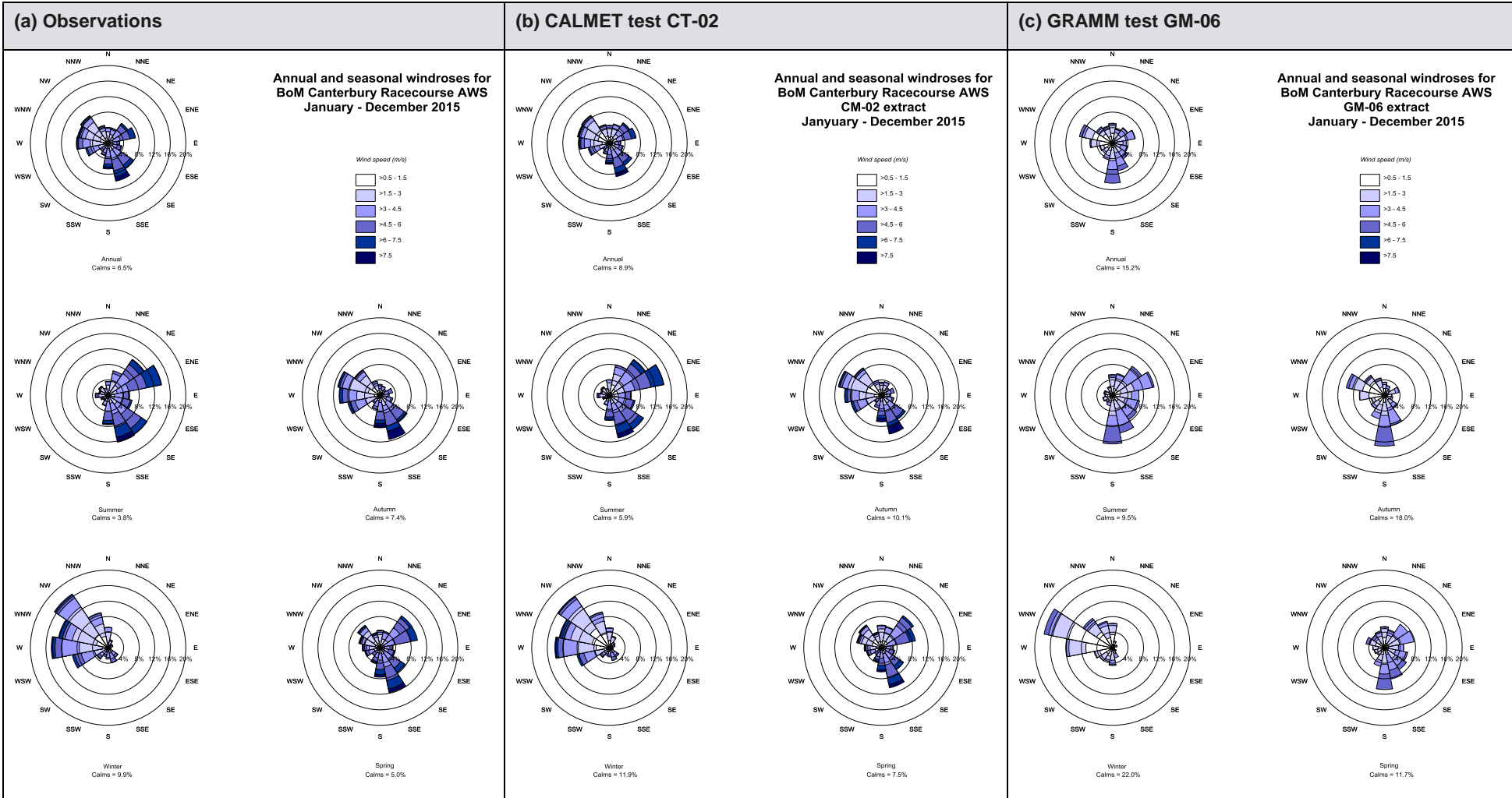
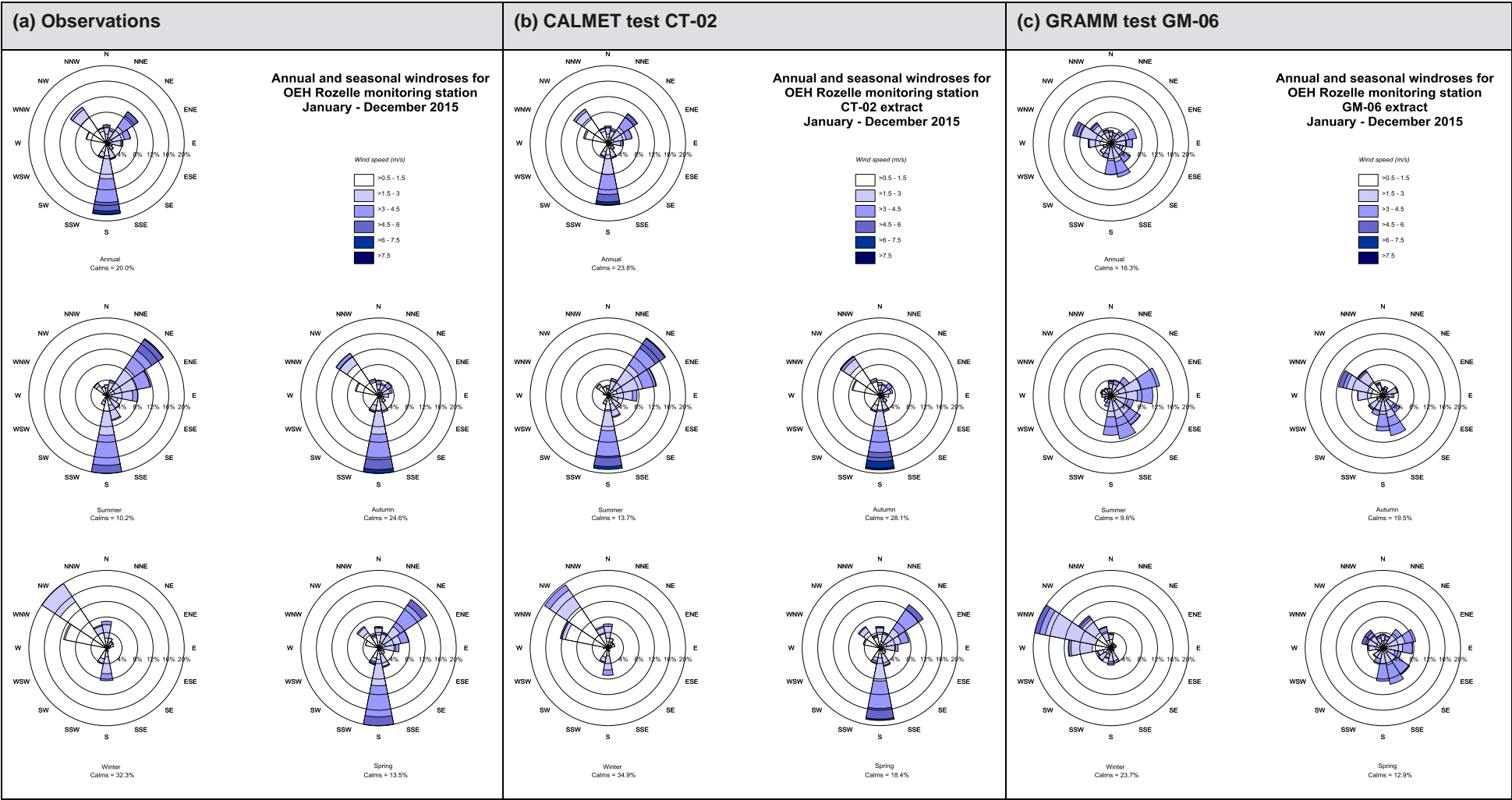
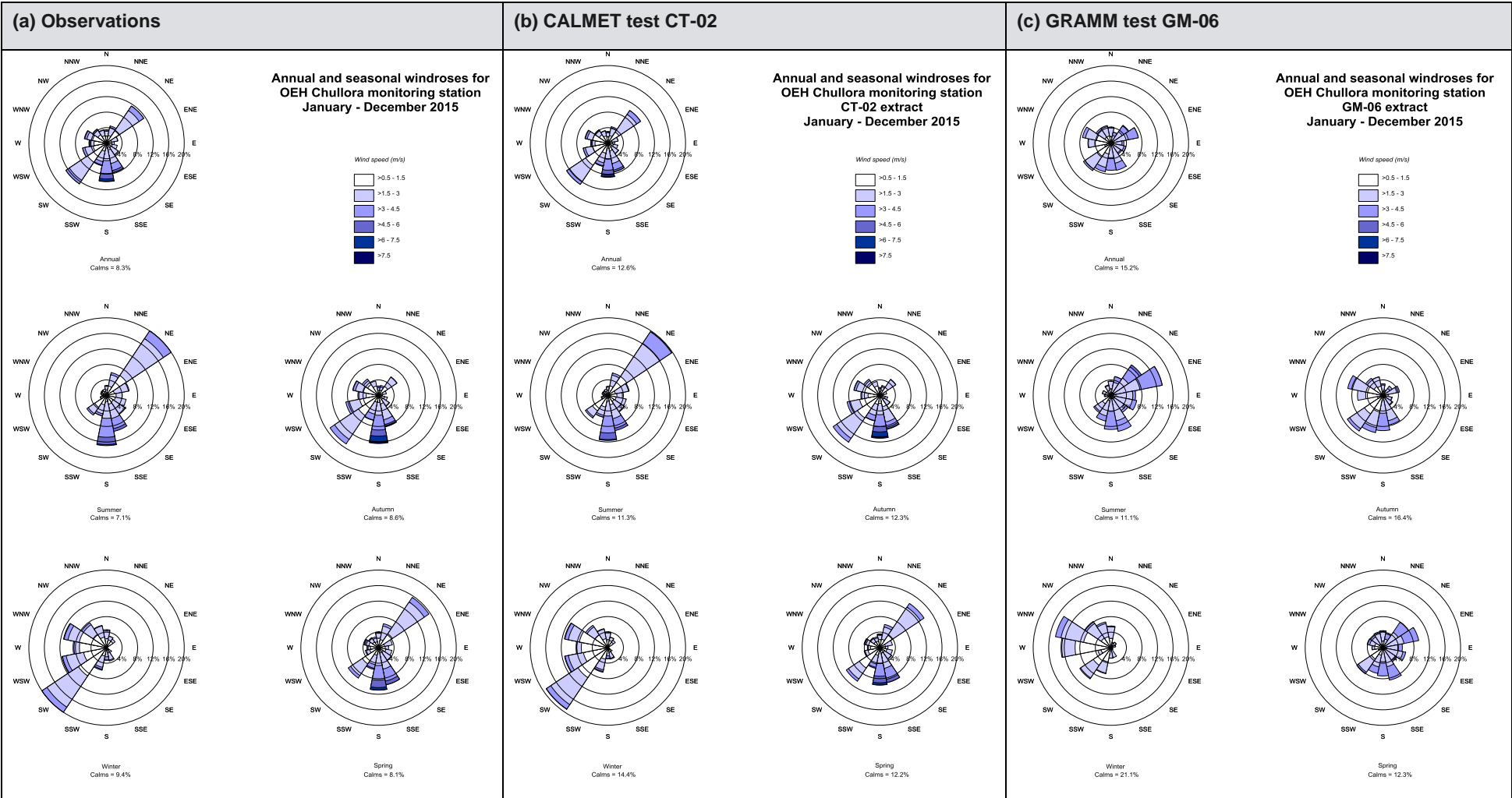


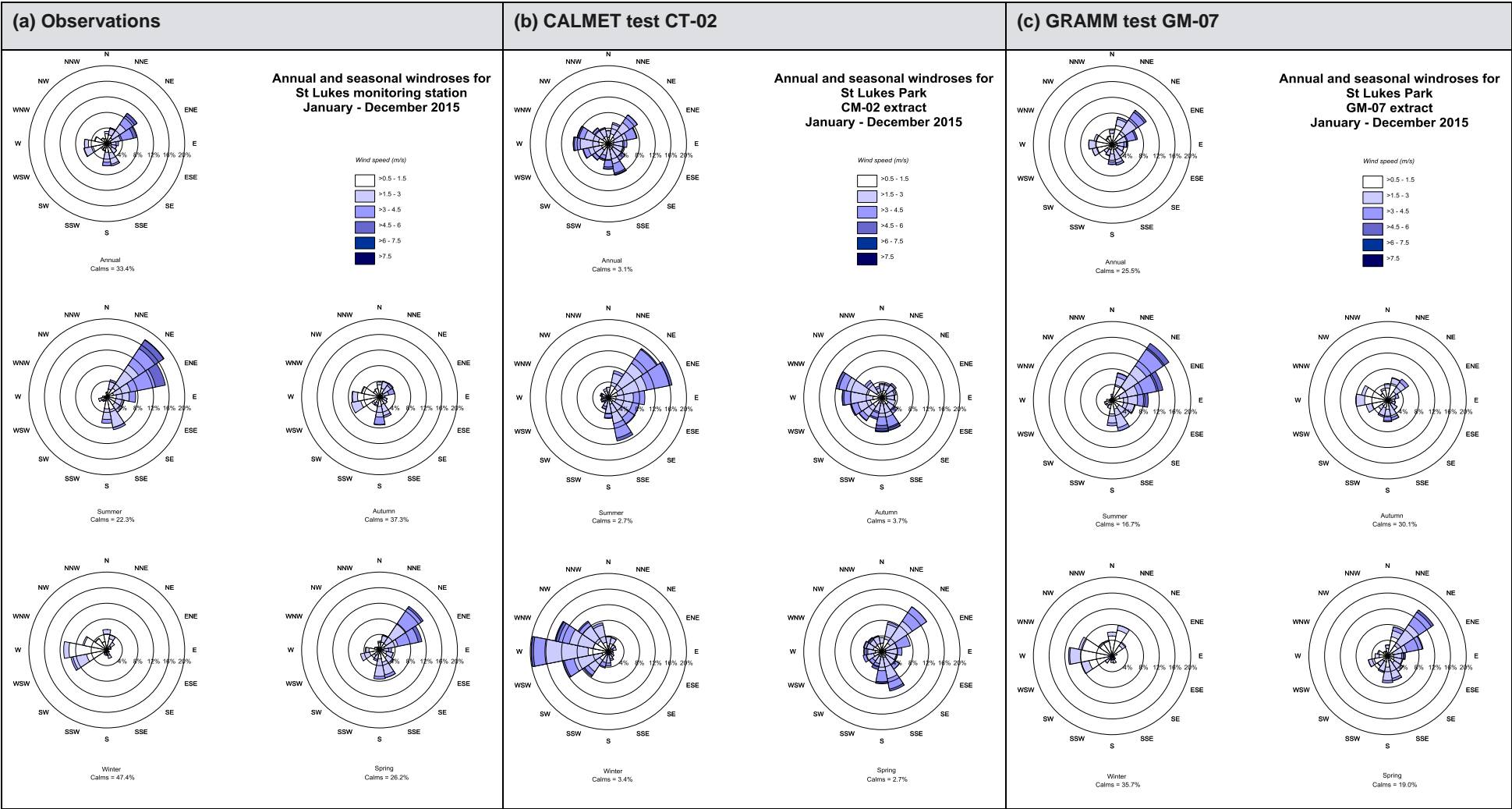
Figure E-55: Model performance - St Lukes Park extract (observations, test CT-02 and test GM-06)











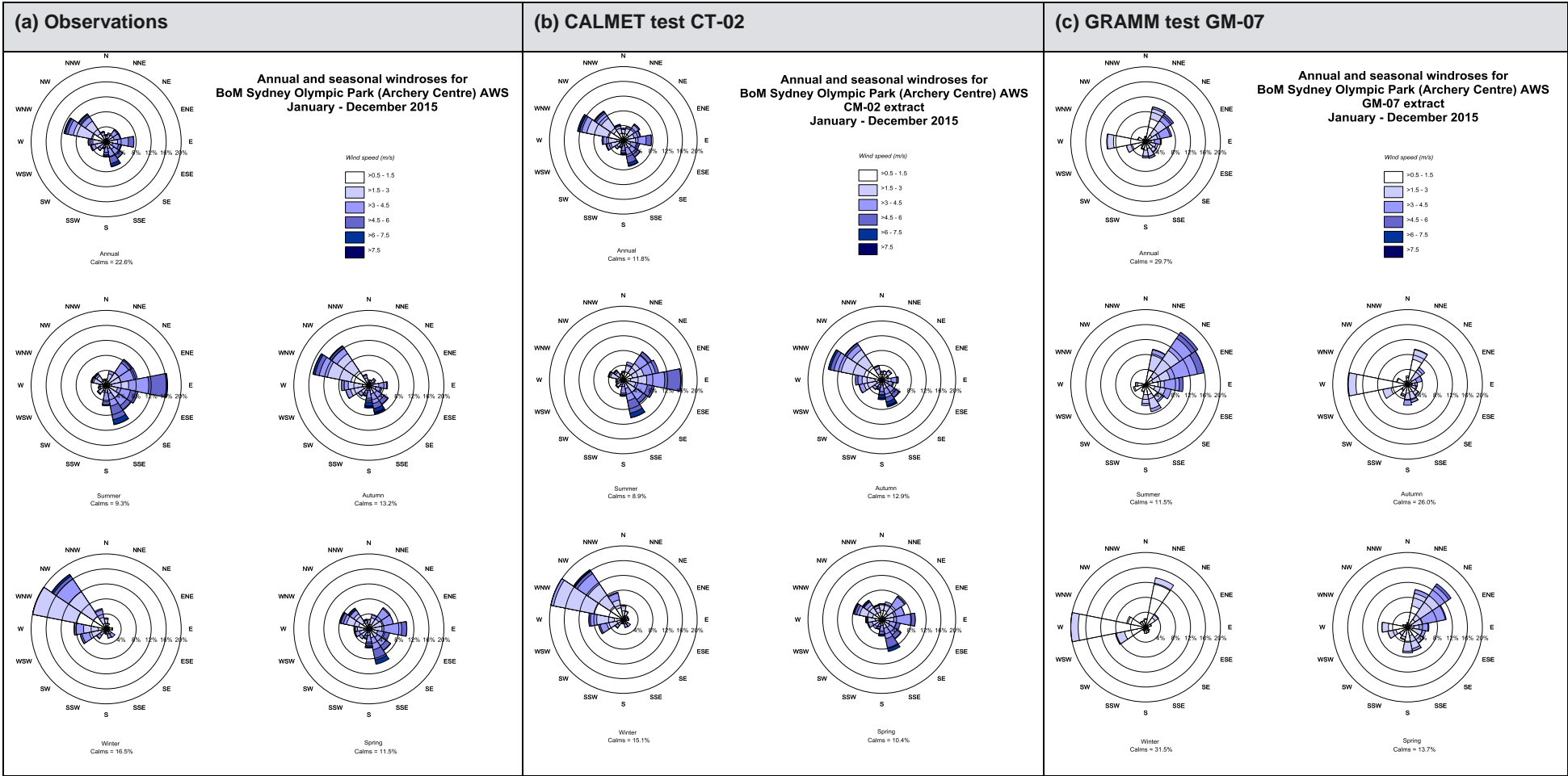


Figure E-61: Model performance – Sydney Olympic Park extract (observations, test CT-02 and test GM-07)

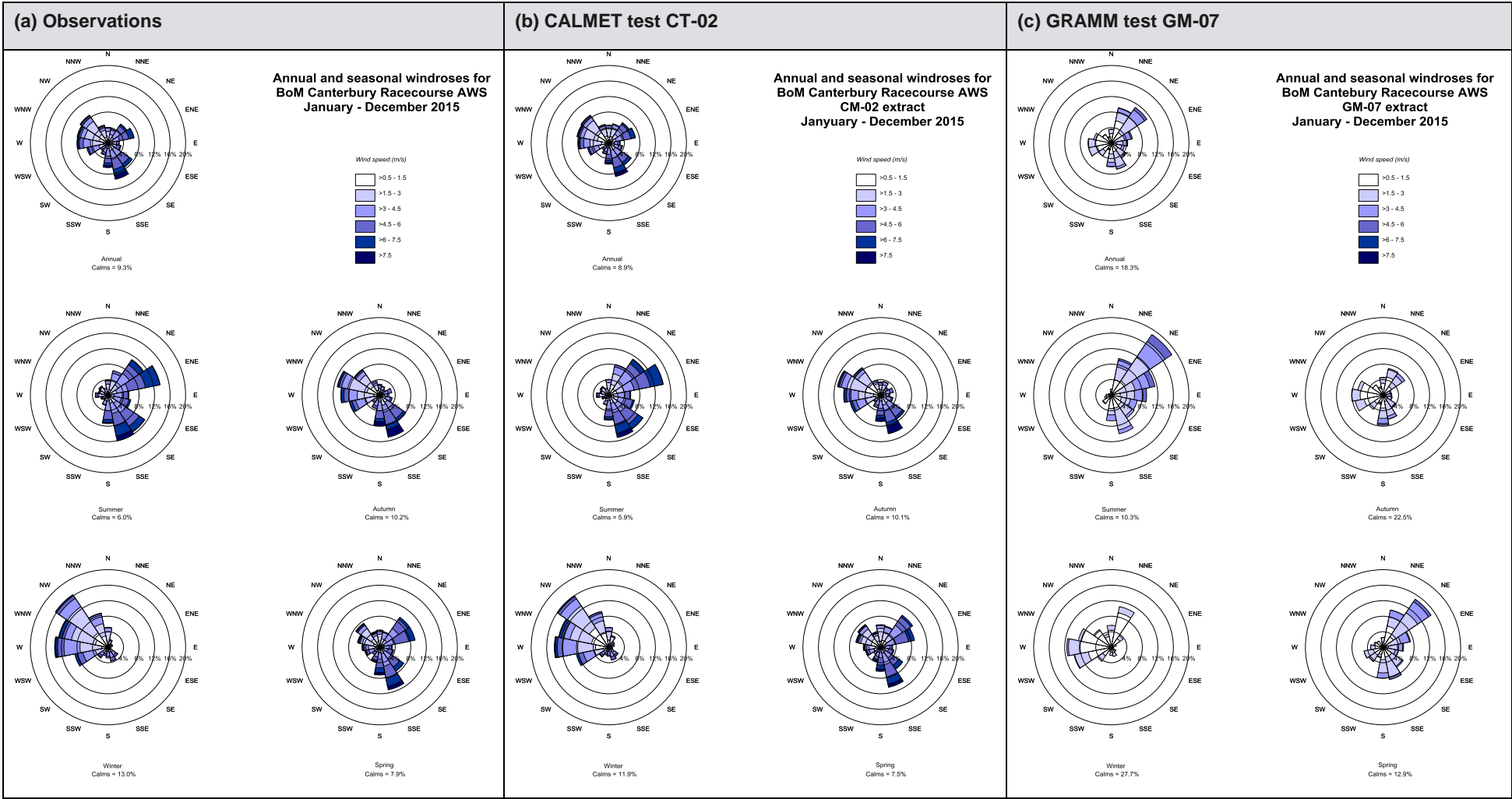
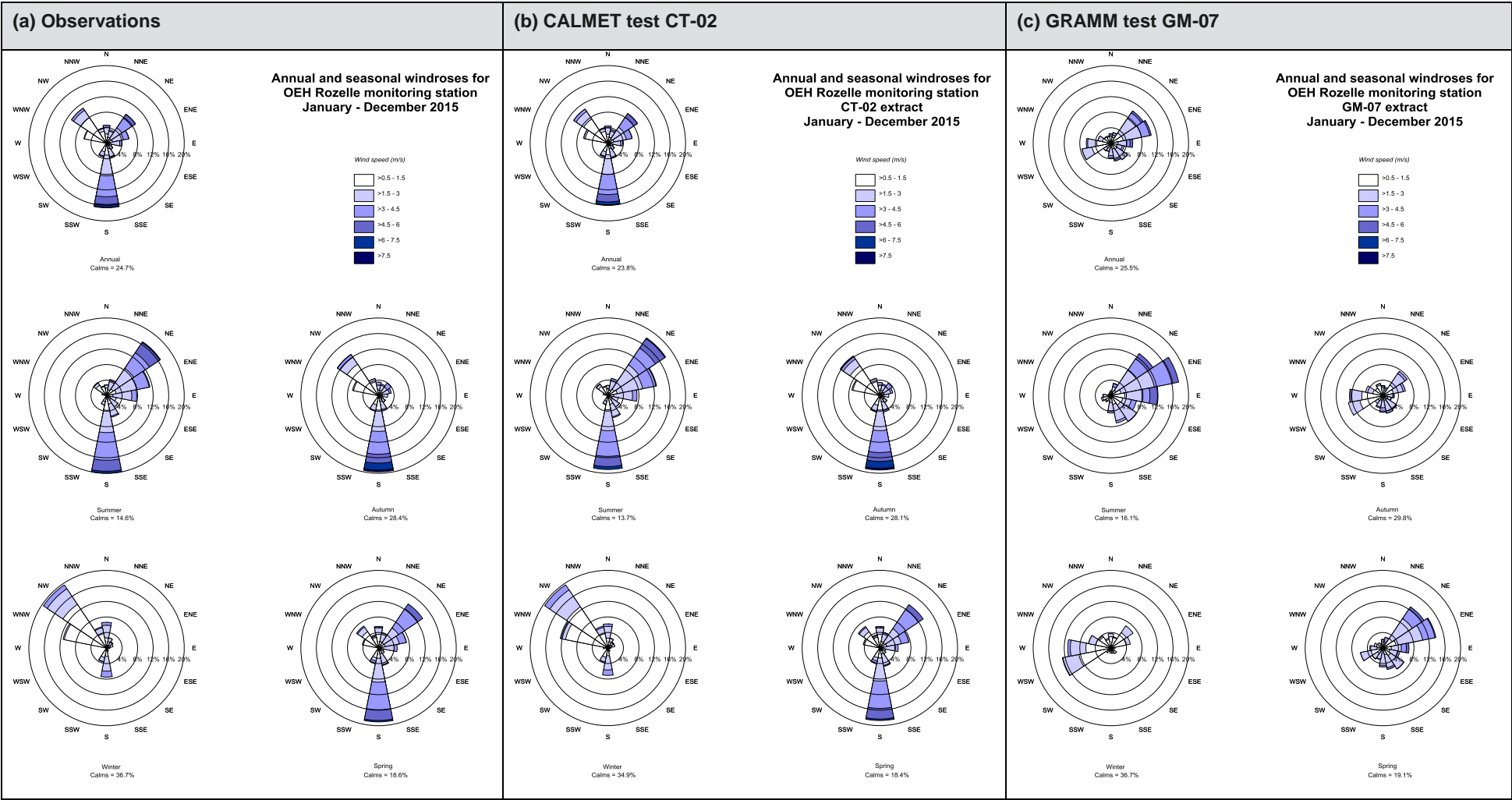


Figure E-62: Model performance – Canterbury Racecourse extract (observations, test CT-02 and test GM-07)



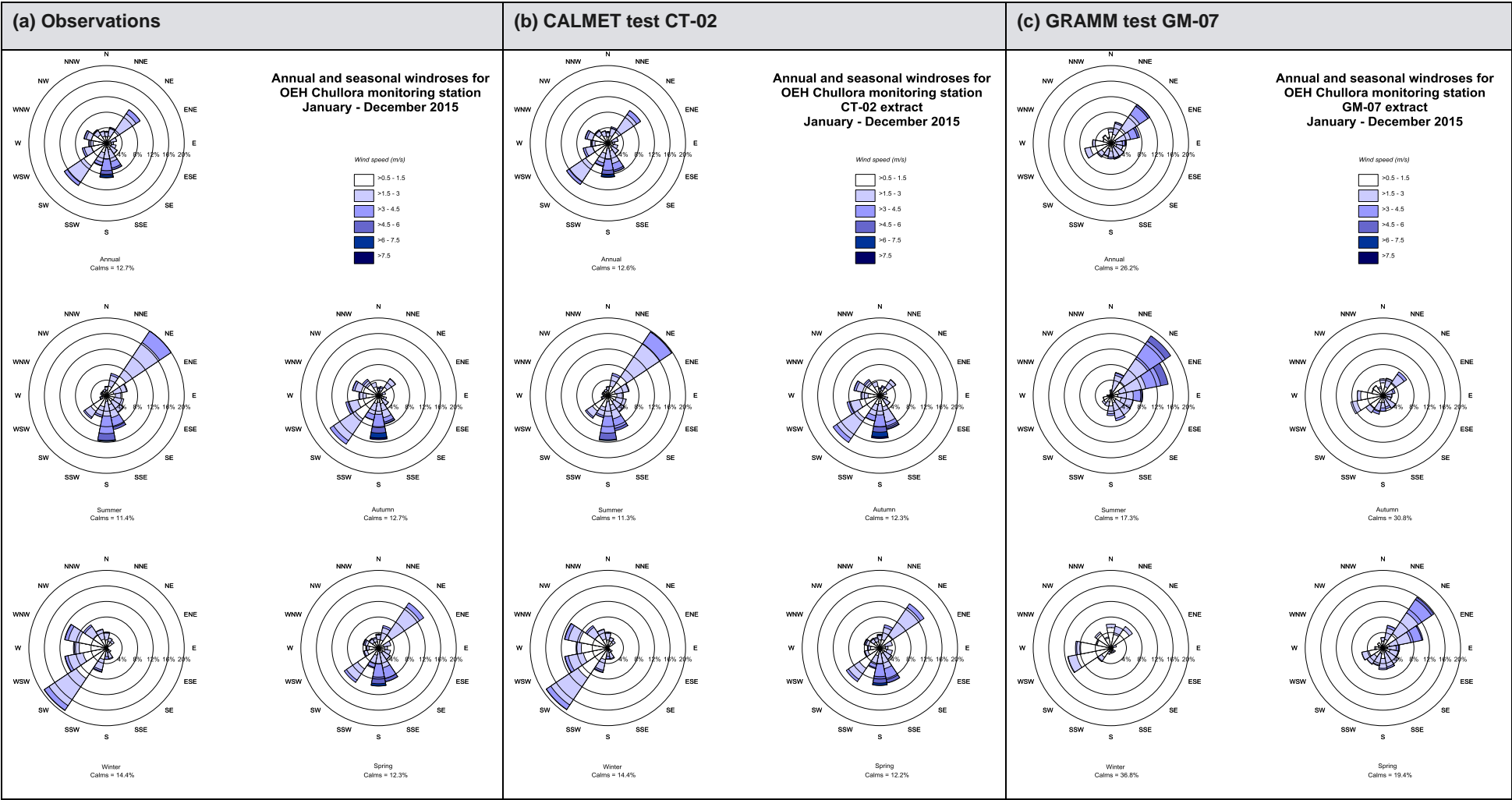


Figure E-64: Model performance – Chullora extract (observations, test CT-02 and test GM-07)

Appendix F

Results of dispersion model evaluation

F1 Temporal analysis (NO_x)

In the temporal analysis of model performance the NO_x concentrations predicted by CAL3QHCR and GRAL were compared with observations from the St Lukes Park and Concord Oval monitoring stations, and for the period between November 2016 and February 2017. The analysis was based on one-hour average data.

The results of the various tests are presented by test series in the following sections. The presentational format is equivalent to that used in the wind speed evaluation in Section D1 (i.e. Openair `timeVariation` plot, Openair `scatterPlot` with linear regression, quantile-quantile plot, Taylor diagram and statistical metrics from the Openair `modStats` function).

In these tests the modelled NO_x component was added to the background observations from St Lukes Park (i.e. using the 'unadjusted background' approach). Whilst the 'adjusted background' approach naturally gave a perfect fit at St Lukes Park, it did not significantly improve the predictions at Concord Oval, and therefore the results are not presented in detail in the report.

F1.1 Series C: CAL3QHCR vs GRAL (tests C3-01 and GL-01)

The results for tests C3-01 (CAL3QHCR) and GL-01 (GRAL) are shown for Concord Oval in Figure F-1 and for St Lukes Park in Figure F-2. In these tests the Concord Oval meteorology was used directly in the models; GRAMM was not used in GRAL. To enable the datasets to be distinguished more clearly in the `timeVariation` plots, confidence intervals are not included.

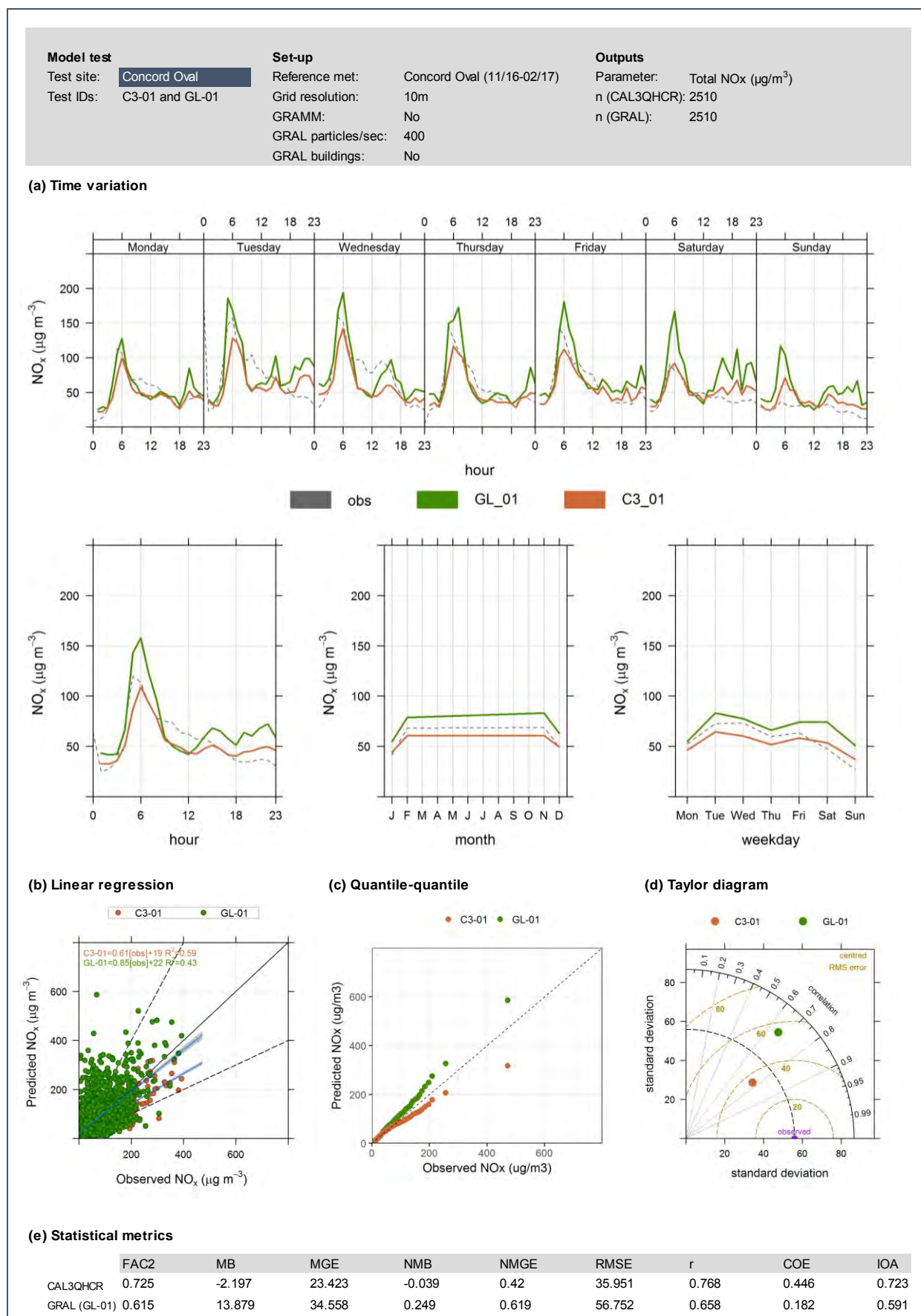
It is helpful to firstly consider the results for the St Lukes Park background site in Figure F-2. Because the model predictions were added to the observations from St Lukes Park, the predicted total concentration was always larger than or equal to the observed concentration. However, as the modelled contribution was generally small at this location, the relationship between the predictions and the observations was naturally very good. The over-prediction was slightly higher on average using GRAL than using CAL3QHCR. For a small proportion of hours GRAL significantly over-estimated the concentration at St Lukes Park.

For the Concord Oval site, the linear regression plot in Figure F-1 shows that, at this high temporal resolution, the performance of both CAL3QHCR and GRAL was not particularly good. The R² value for CAL3QHCR (0.59) was higher than that for GRAL (0.43), but on the other hand CAL3QHCR resulted in a general under-prediction of NO_x (by around 35% for an observed concentration of 400 µg/m³), whereas GRAL was more accurate overall.

The simulation of the average temporal patterns in the NO_x concentrations at Concord Oval was much better, with a reasonably accurate representation of diurnal patterns. GRAL tended to overestimate concentrations during the morning peak traffic period, whereas CAL3QHCR tended to underestimate these. Because of the use of an average weekday emissions profile in the modelling, both models overestimated NO_x concentrations on Saturdays and Sundays. Although the removal of the results for weekends improved the predictions, the improvement was not very large overall (at Concord Oval the R² for CAL3QHCR increased from 0.59 to 0.61, and for GRAL it increased from 0.43 to 0.48).

The Taylor diagram shows that GRAL tended to overestimate the variation in the observations, and CAL3QHCR tended to underestimate it.

On balance, and from an air quality assessment point of view, the slight over-estimations of concentrations in GRAL would be preferable to the slight underestimation in CAL3QHCR.

Figure F-1: Dispersion model performance for NO_x – Concord Oval (tests C3-01 and GL-01)

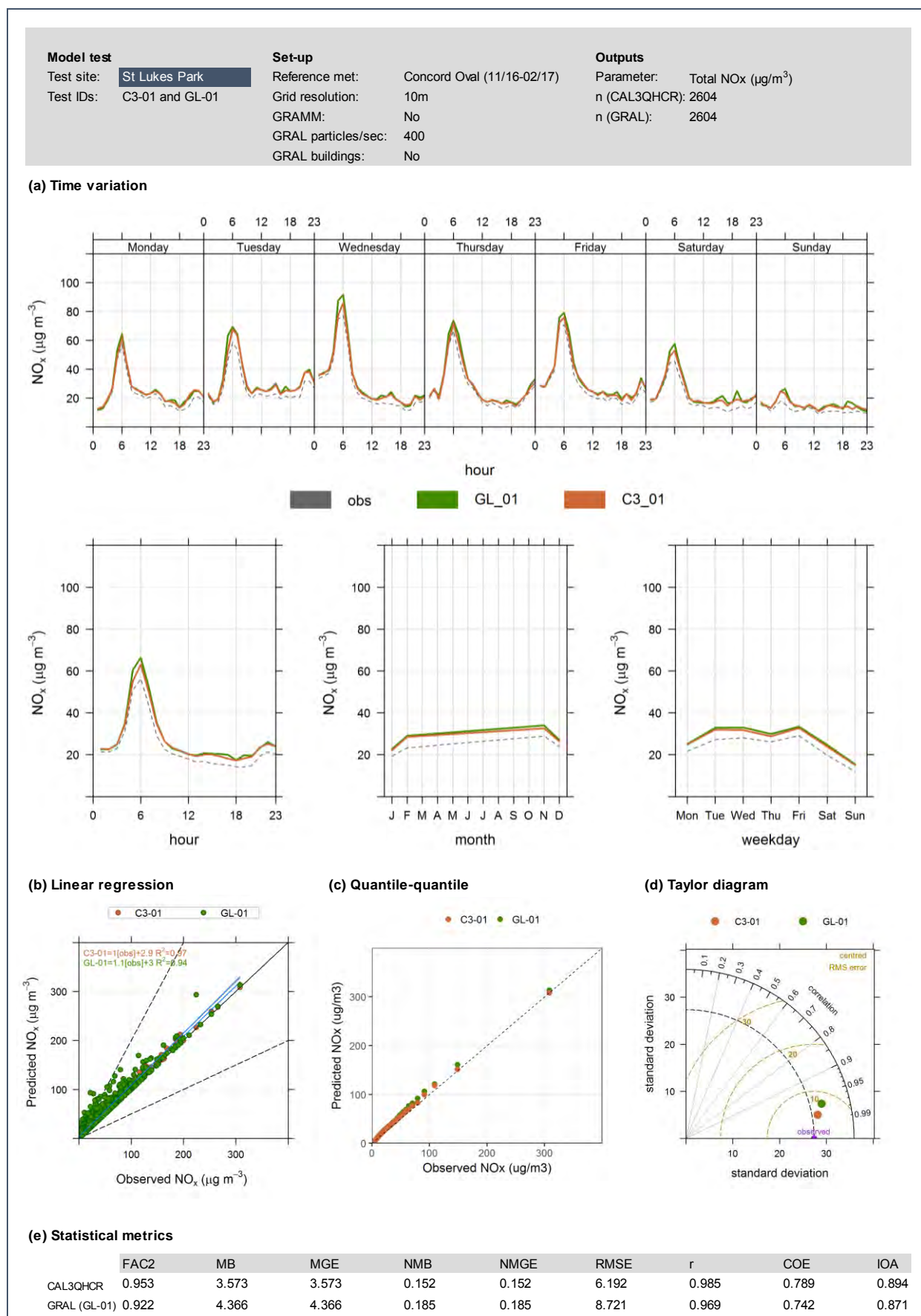


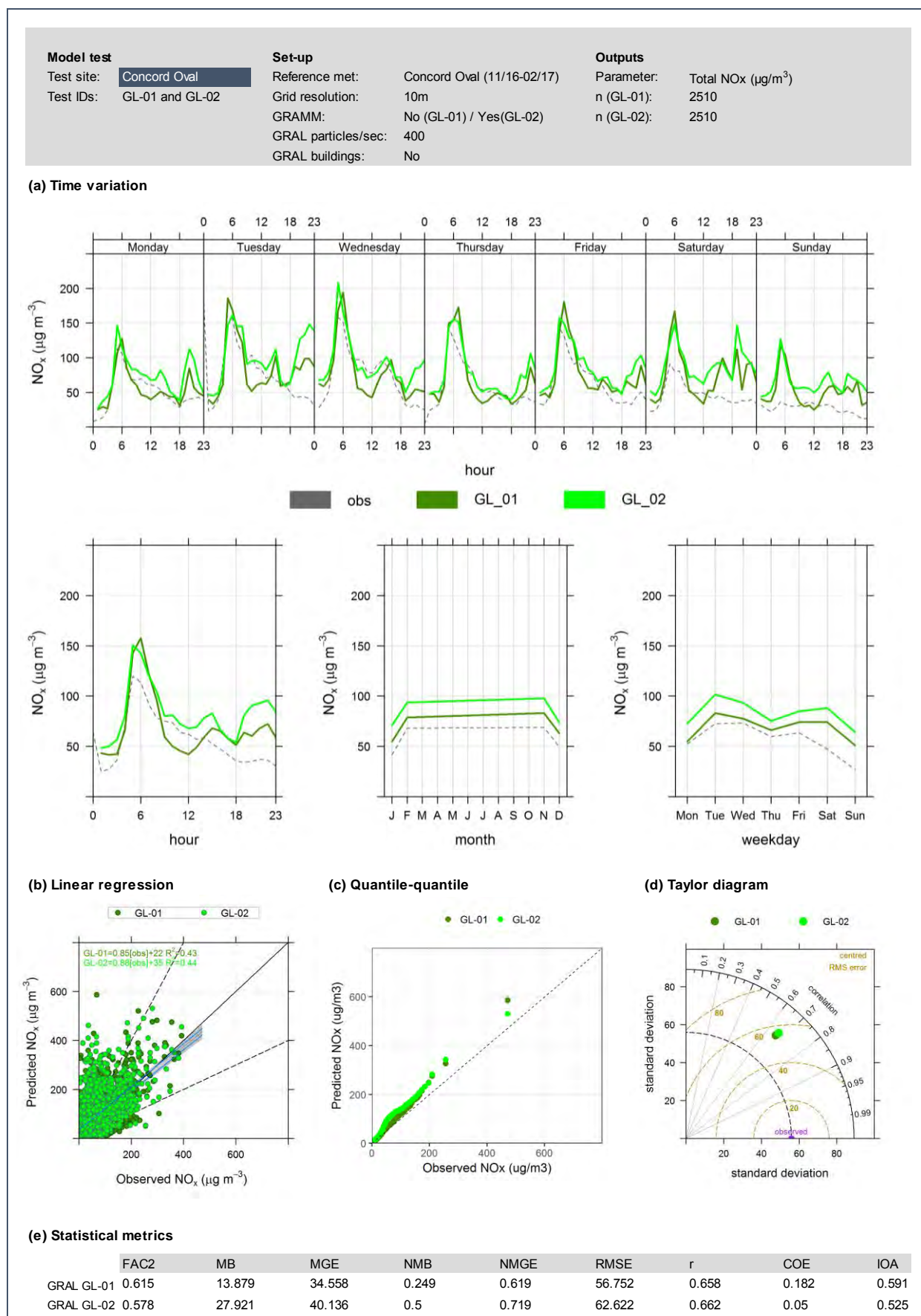
Figure F-2: Dispersion model performance for NO_x – St Lukes Park (tests C3-01 and GL-01)

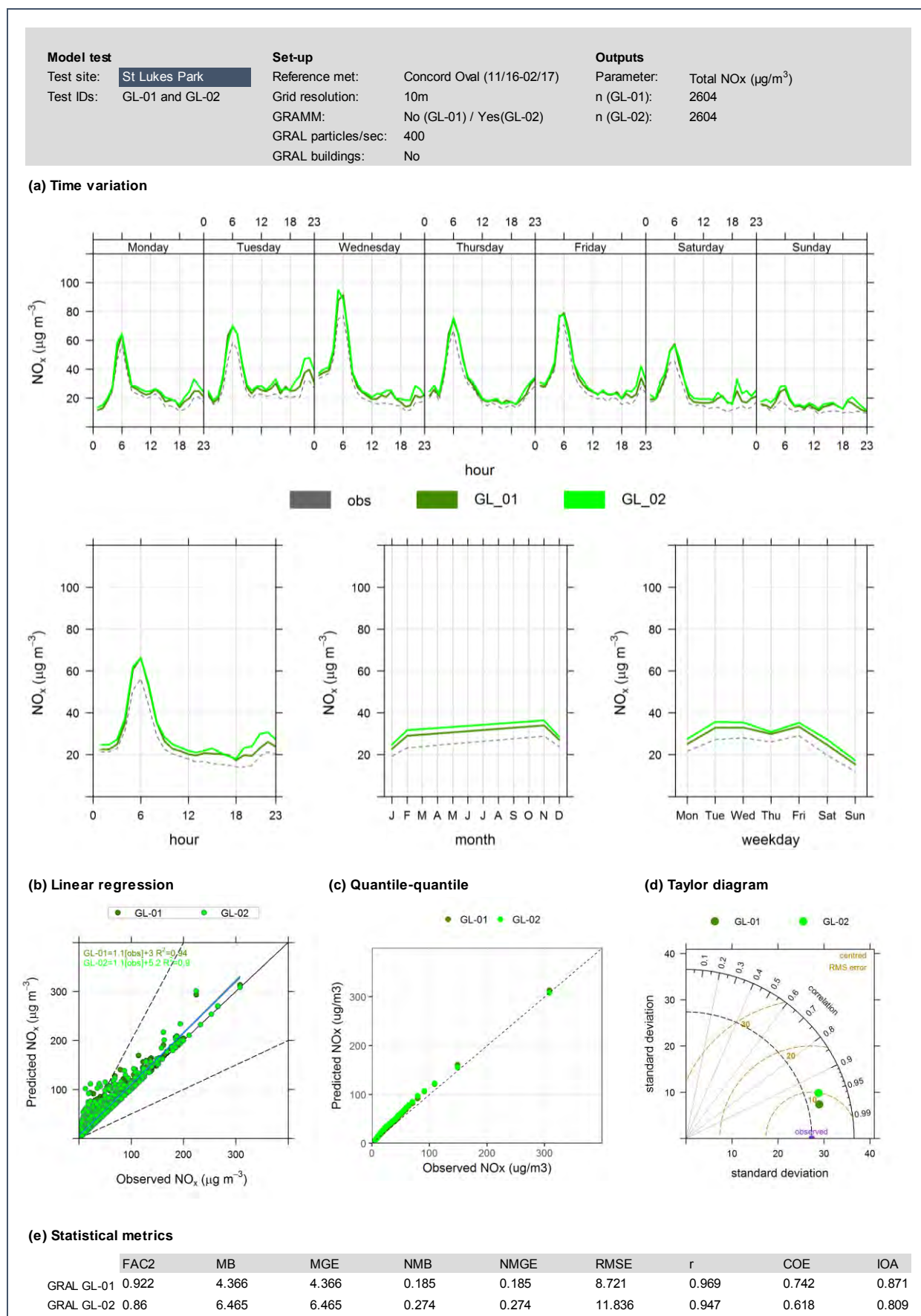
F1.2 Series D: GRAL meteorological input (tests GL-01 and GL-02)

The tests in Series D examined the effects of the meteorological input in GRAL. In test GL-02 GRAMM (Concord Oval Match-to-Observations) was used rather than the direct observations from Concord Oval in test GL-01. Other model settings were not changed.

The results of the tests GL-01 and GL-02 are shown in Figures F-3 and F-4. There were only quite small differences between the GRAL predictions in the two tests. On average, the predictions in test GL-02 were higher than those in test GL-01, especially in the afternoon and evening. During the morning peak period the predictions in both tests were very similar.

Most of statistical metrics indicated a slight deterioration in overall model performance when GRAMM was used.

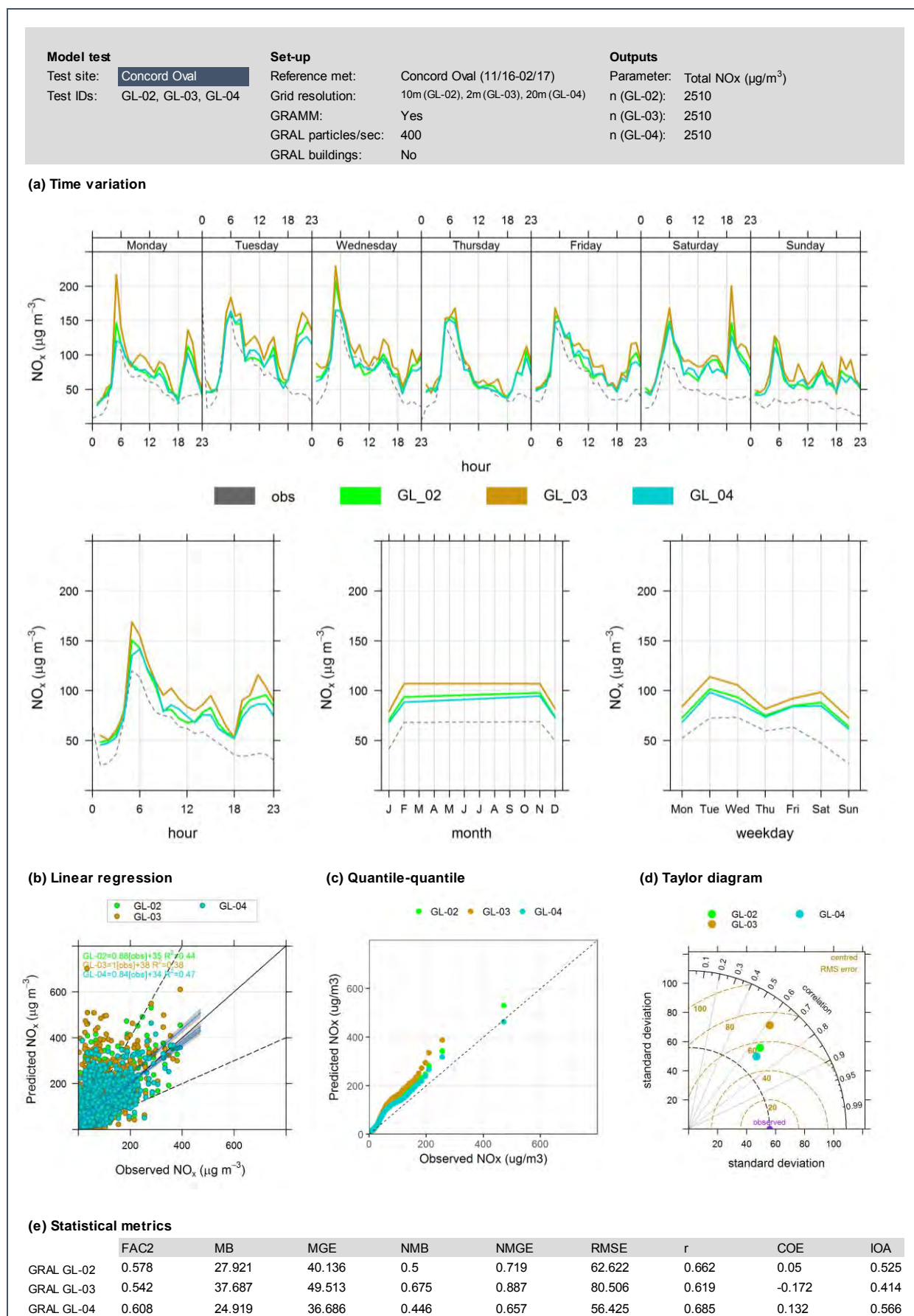
Figure F-3: Dispersion model performance for NO_x – Concord Oval (tests GL-01 and GL-02)

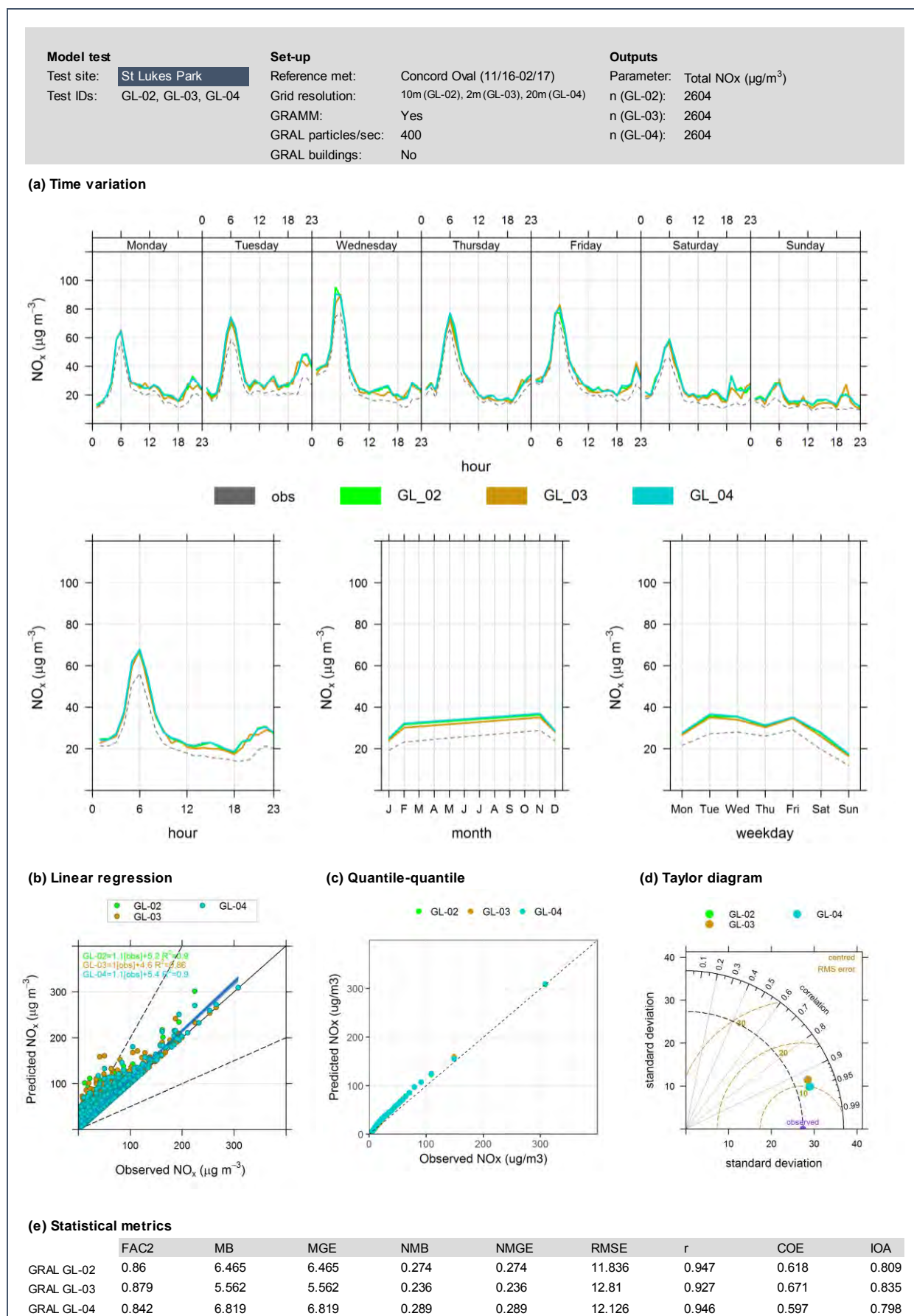
Figure F-4: Dispersion model performance for NO_x – St Lukes Park (tests GL-01 and GL-02)

F1.3 Series E: GRAL grid spacing (tests GL-02, GL-03 and GL-04)

The Series E tests examined the influence of the grid spacing in GRAL. The number of particles in GRAL was fixed at 400 per second. The results for the different grid spacing values in GRAL (2, 10 and 20 metres) are shown in Figure F-5 and F-6.

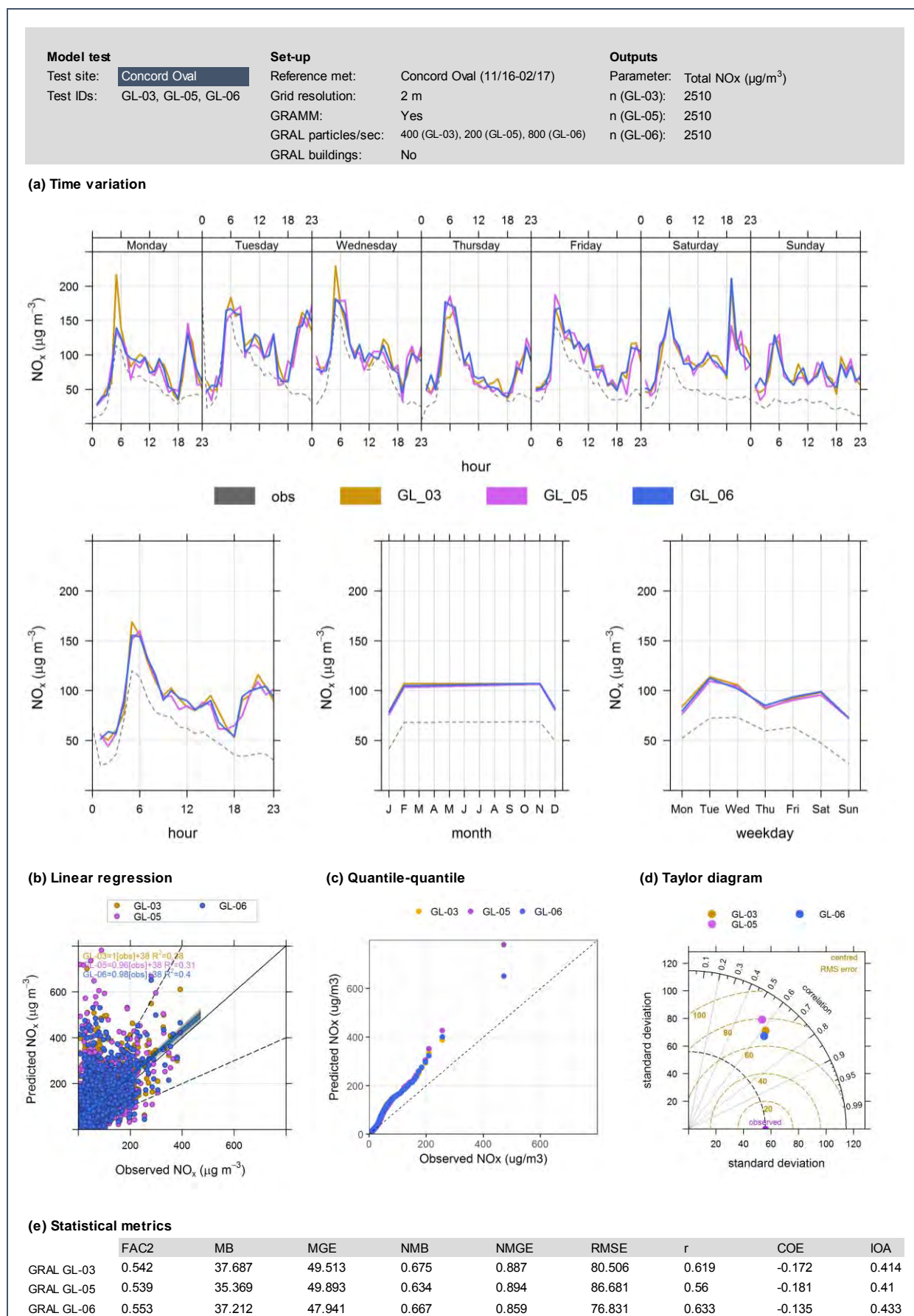
At Concord Oval, the predictions for the three grid spacings were quite similar. The highest concentrations were obtained using the finest resolution, and the lowest concentrations with the coarsest resolution. Overall the test using the coarsest resolution actually resulted in the closest agreement with the predictions. This observation is not surprising as concentration gradients are generally steep close to line sources and therefore a coarser grid usually results in lower values. As the GRAL results have shown a general overestimation overall, it therefore makes sense that, in this study, the coarser resolution test showed results closest to the measurements.

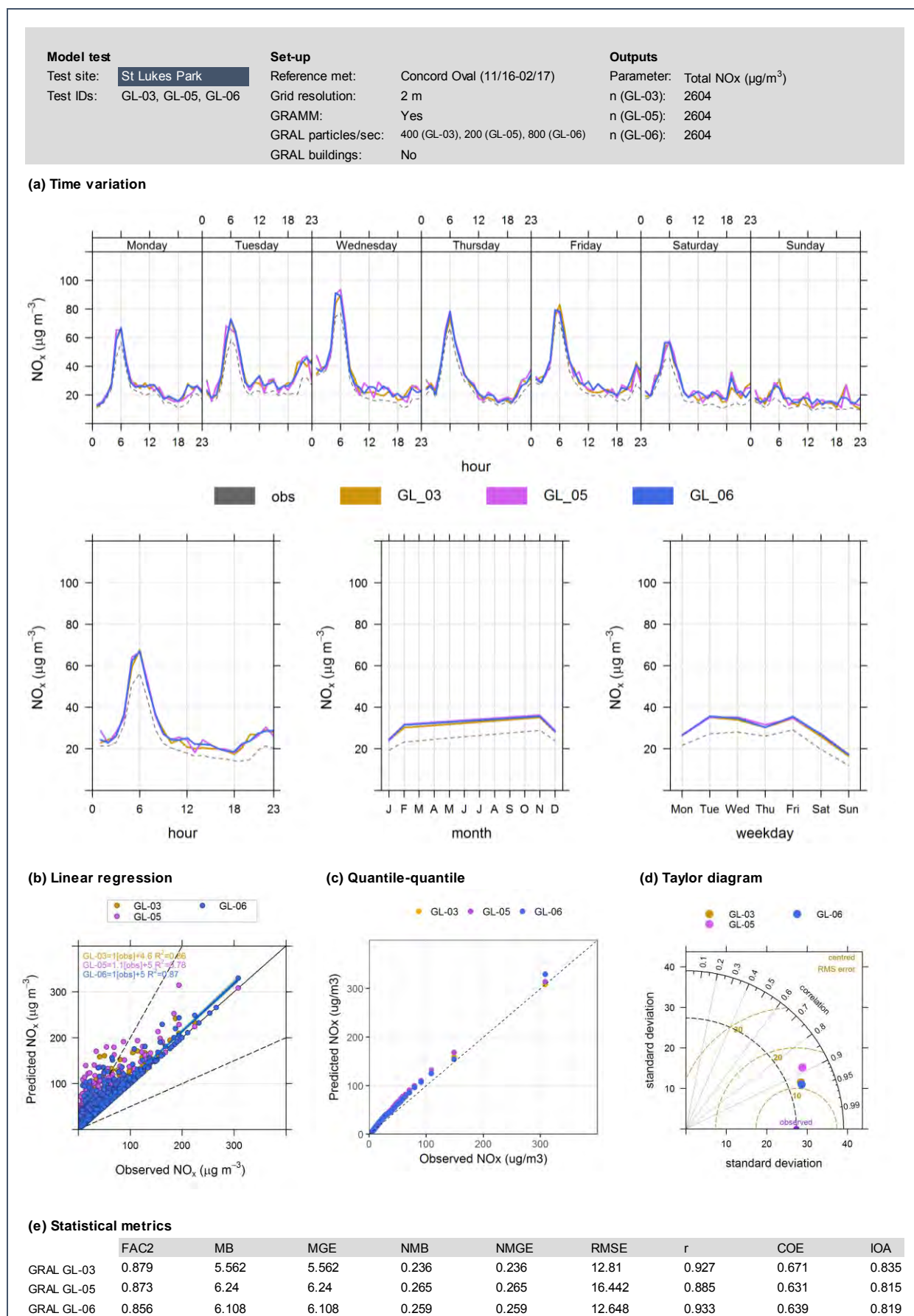
Figure F-5: Dispersion model performance for NO_x – Concord Oval (tests GL-02, GL-03 and GL-04)

Figure F-6: Dispersion model performance for NO_x – St Lukes Park (tests GL-02, GL-03 and GL-04)

F1.4 Series F: GRAL particle number (tests GL-03, GL-05 and GL-06)

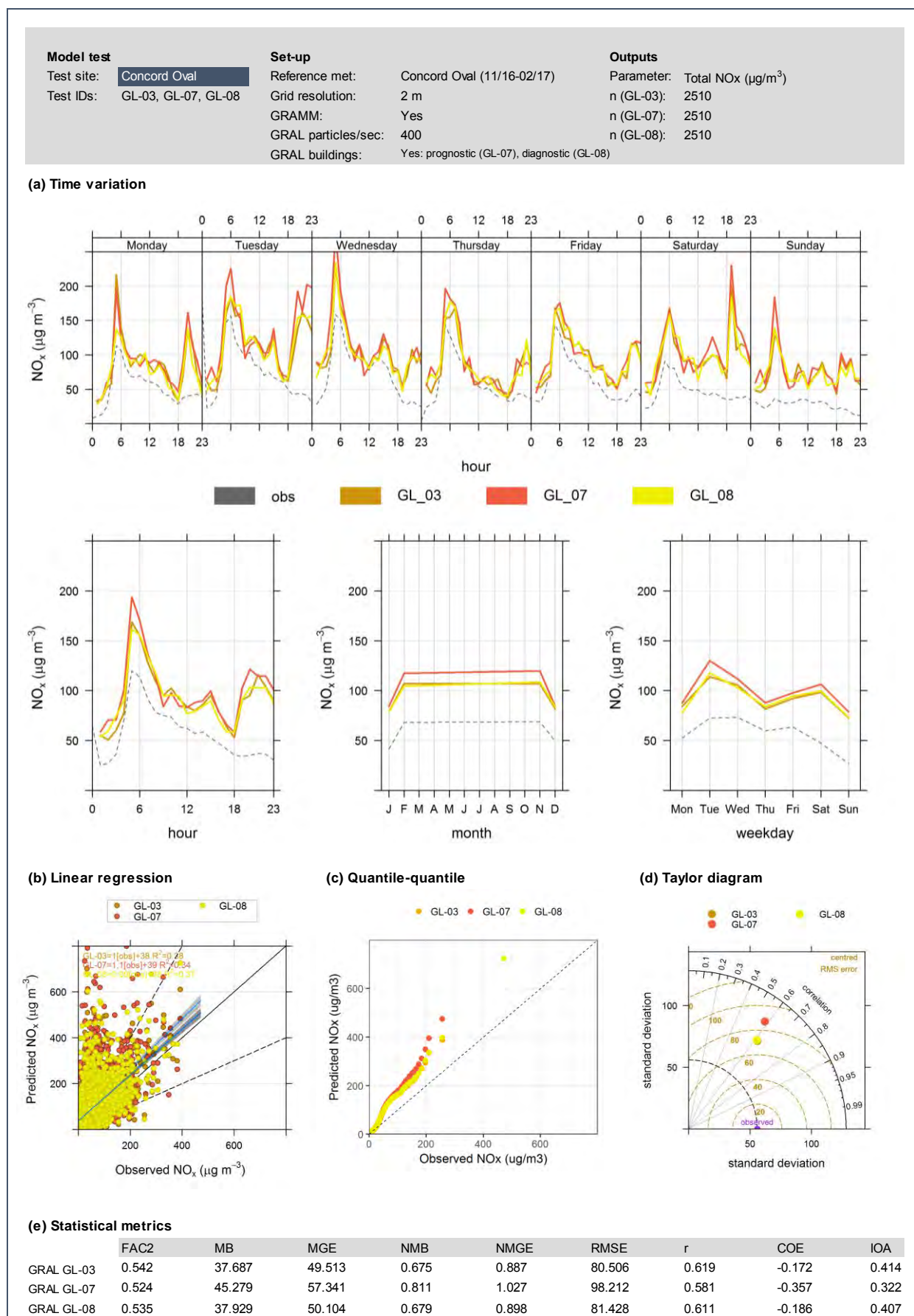
Figures F-7 and F-8 show that the number of particles used in GRAL (200, 400 or 800) had little effect on the model predictions. The only noticeable difference between the test results was that test GL-03 (400 particles per second) gave markedly higher morning peak concentrations on Mondays and Wednesdays than tests GL-05 or GL-06. The reason for this is not clear. However, it could simply be a result of model iterations causing particles to stay in a particular cell for longer or shorter in one test than in another.

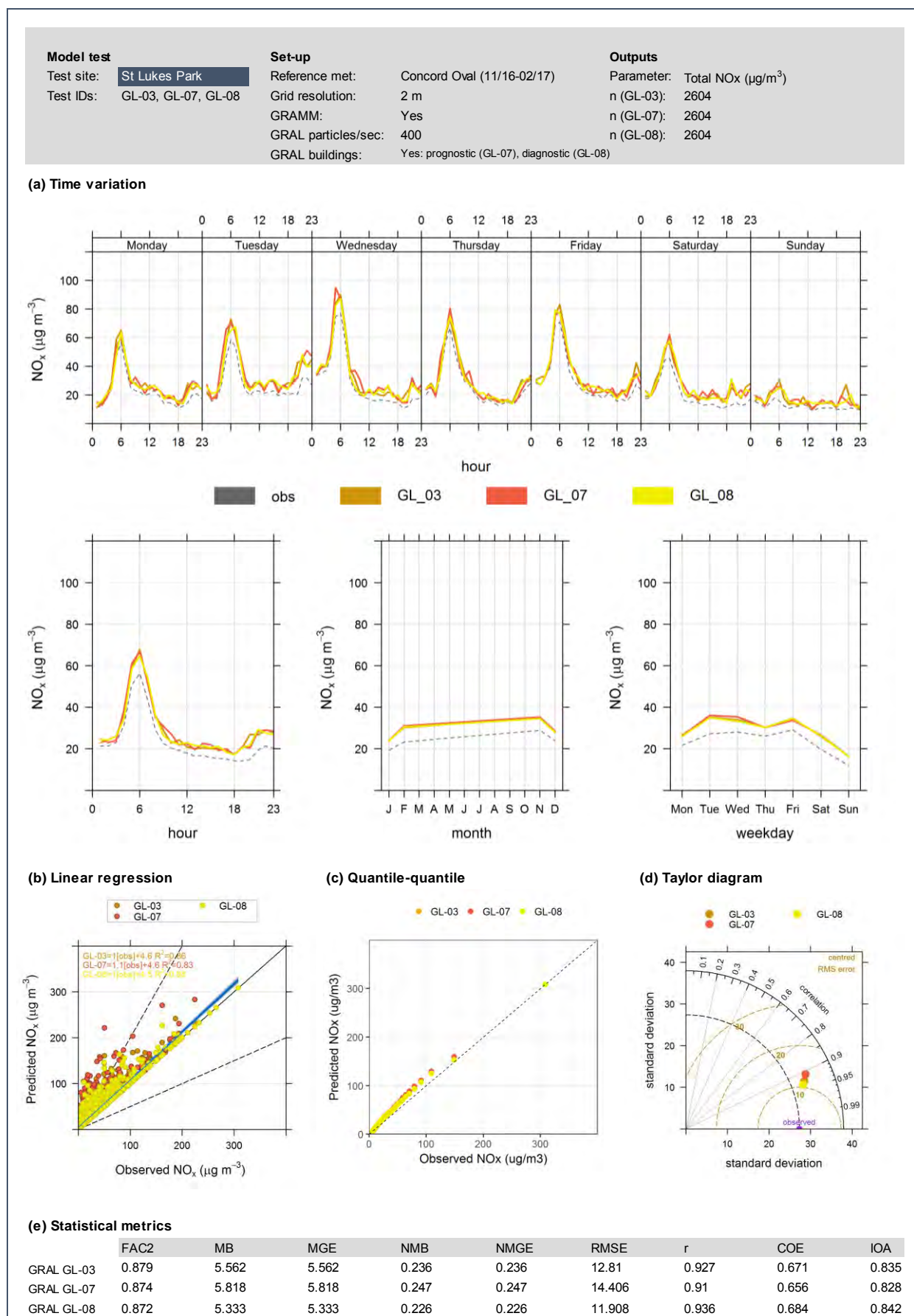
Figure F-7: Dispersion model performance for NO_x – Concord Oval (tests GL-03, GL-05 and GL-06)

Figure F-8: Dispersion model performance for NO_x – St Lukes Park (tests GL-03, GL-05 and GL-06)

F1.5 Series G: GRAL buildings (tests GL-03, GL-07 and GL-08)

The effects of including buildings in GRAL are shown in Figures F-9 and F-10. This included the separate testing of prognostic (GL-07) vs diagnostic (GL-08) approaches. The results were compared with those from test GL-03 (no buildings).

Figure F-9: Dispersion model performance for NO_x – Concord Oval (tests GL-03, GL-07 and GL-08)

Figure F-10: Dispersion model performance for NO_x – St Lukes Park (tests GL-03, GL-07 and GL-08)

F2 Contour plots

This Section contains contour plots for the dispersion model domain. The plots are for the average and maximum 1-hour NO_x concentrations during the four months of the study.

F2.1 Average NO_x concentrations

Figures F-11 to F-21 show the contour plots for average NO_x concentration (model only, no background). For the tests including buildings (GL-07 and GL-08), additional plots are included to provide detail around Concord Oval.



Figure F-11: Contour plot: average NO_x concentration (test C3-01)

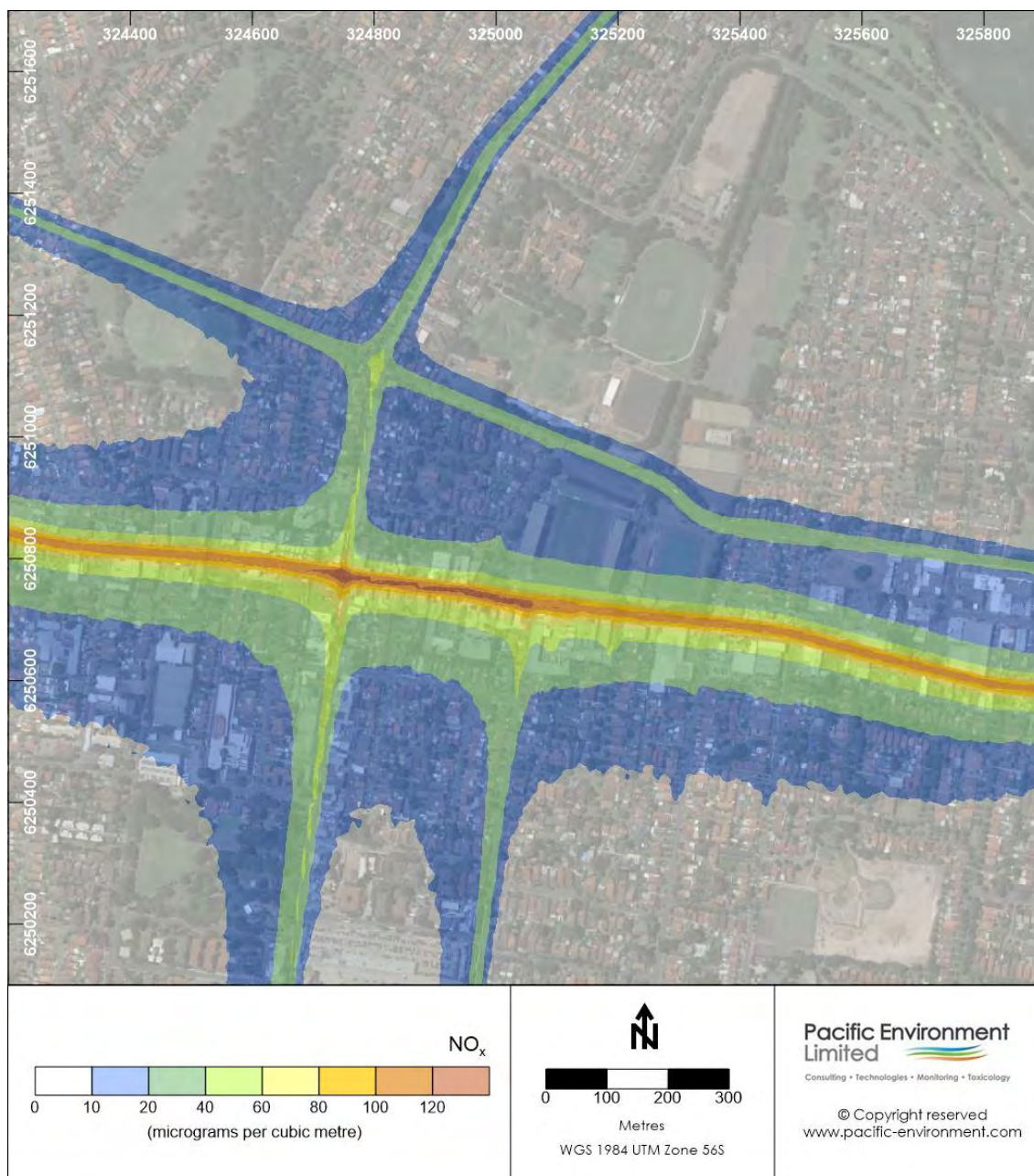


Figure F-12: Contour plot: average NO_x concentration (test GL-01)

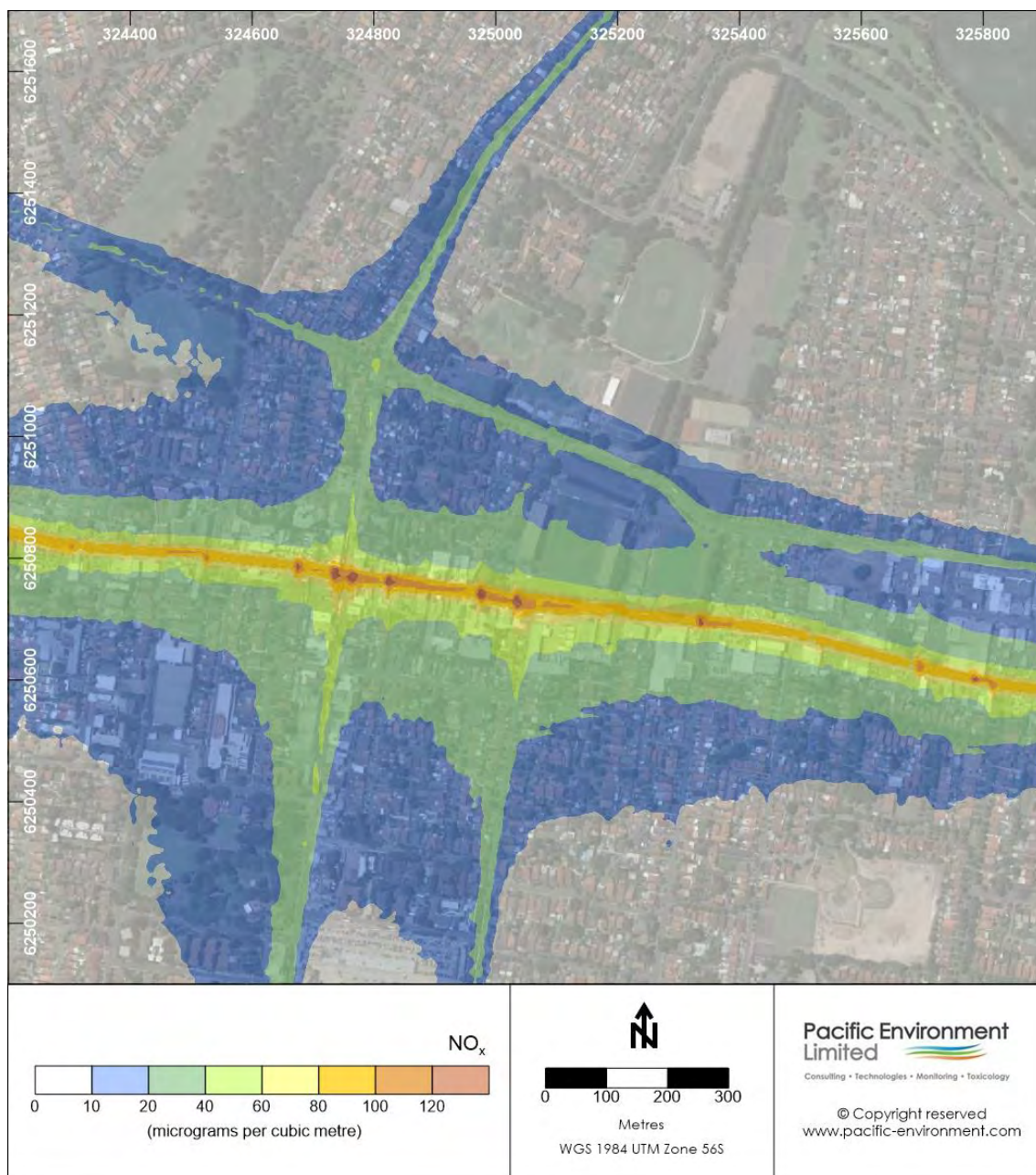


Figure F-13: Contour plot: average NO_x concentration (test GL-02)

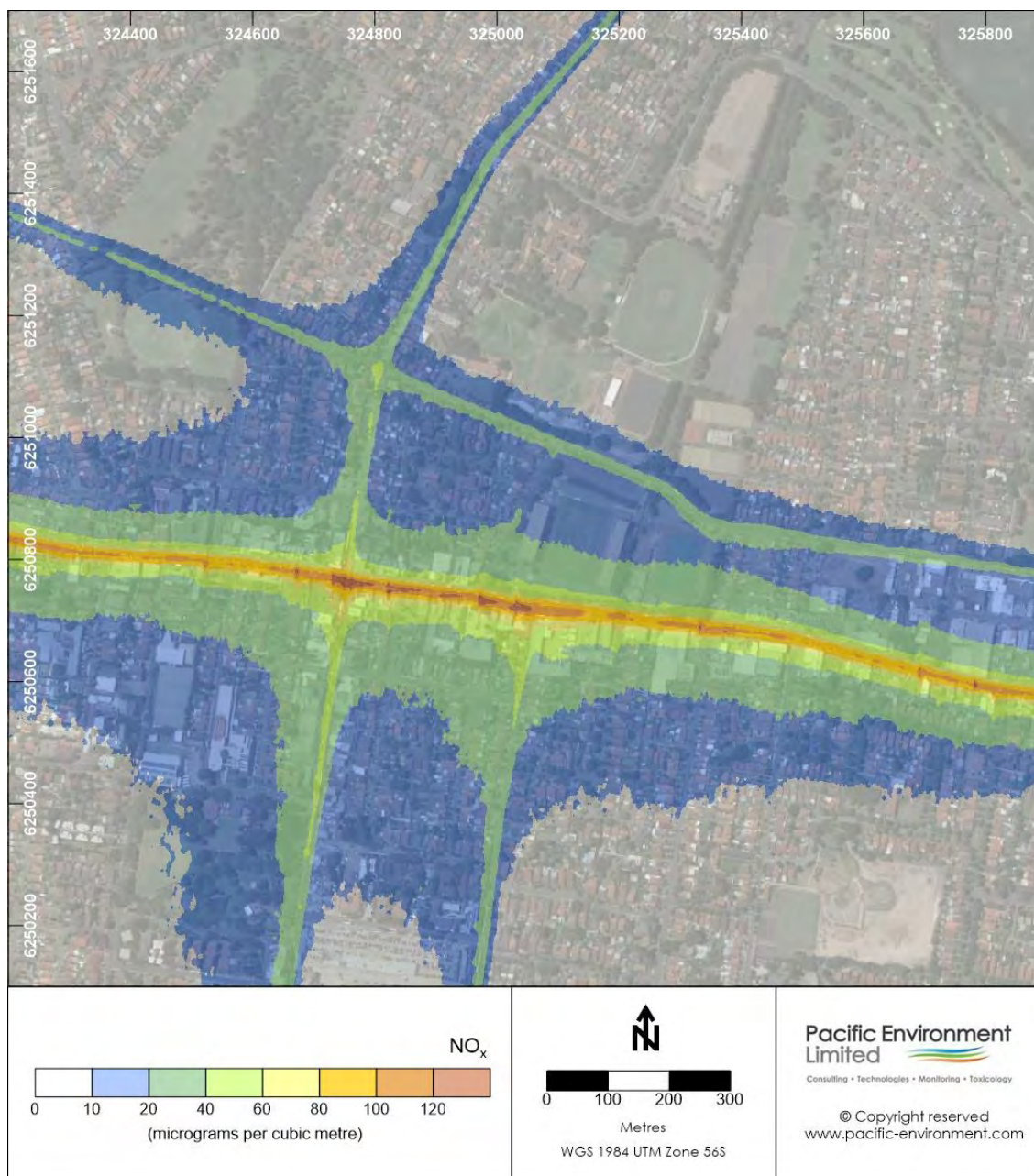


Figure F-14: Contour plot: average NO_x concentration (test GL-03)

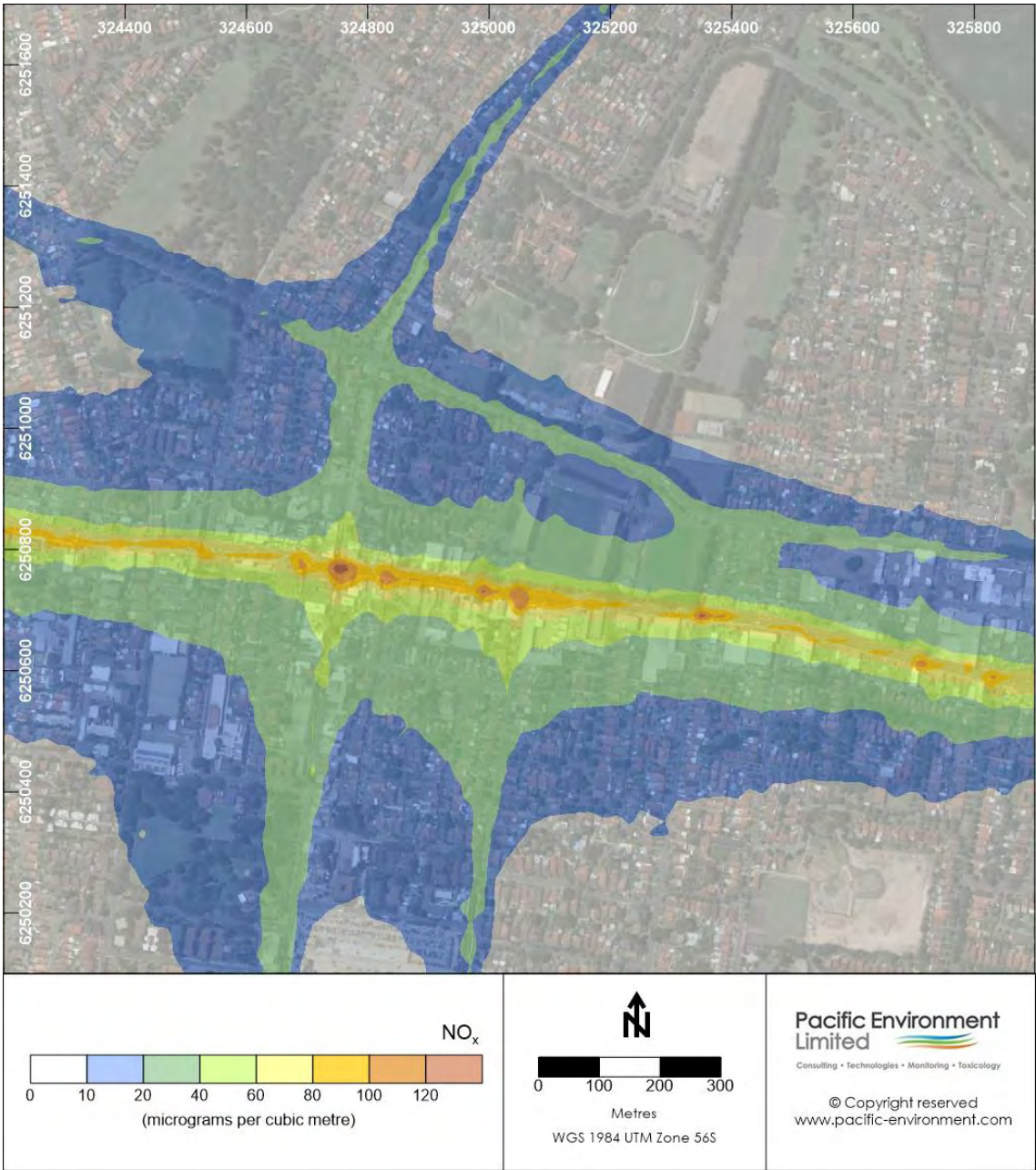


Figure F-15: Contour plot: average NO_x concentration (test GL-04)

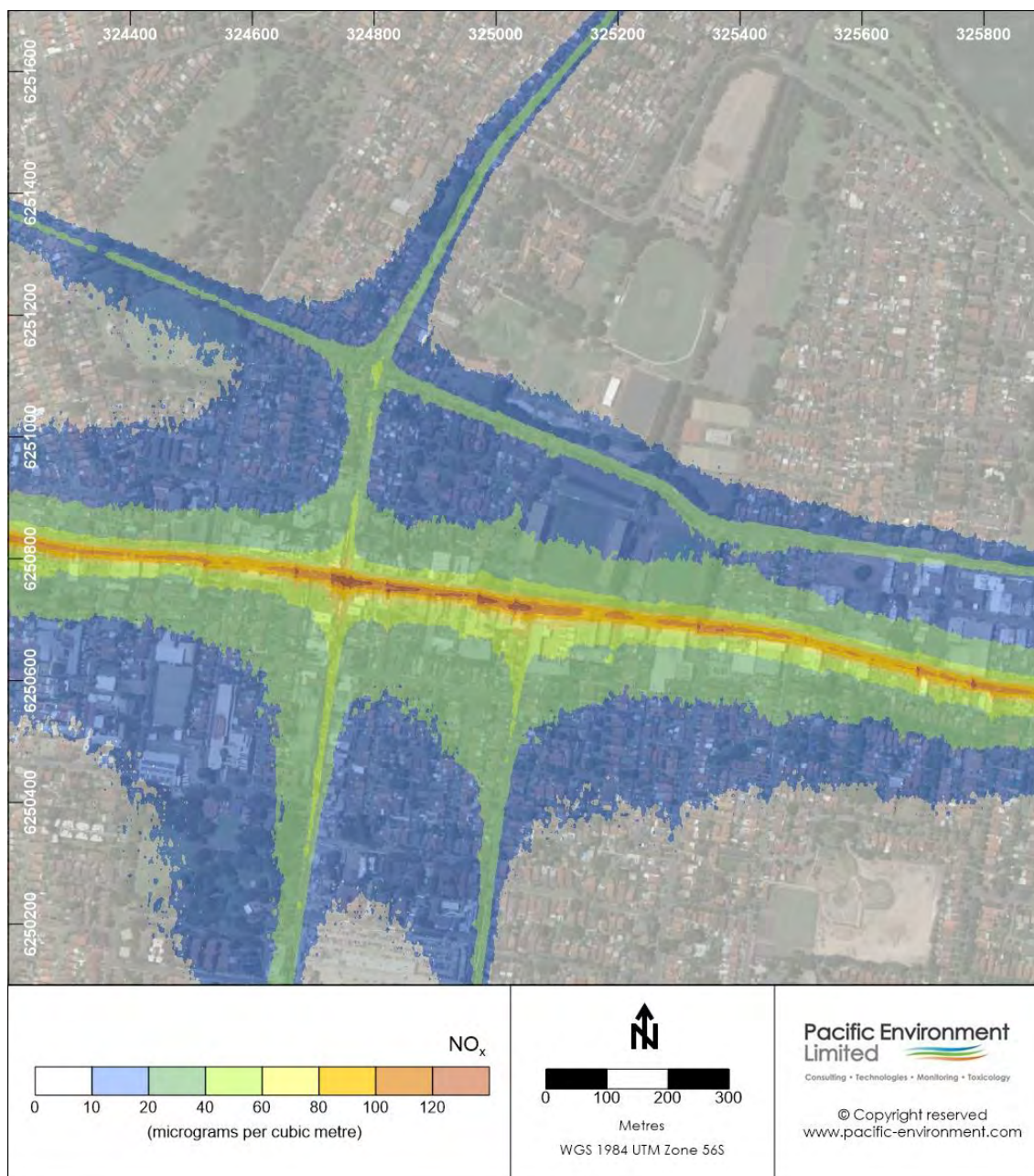


Figure F-16: Contour plot: average NO_x concentration (test GL-05)

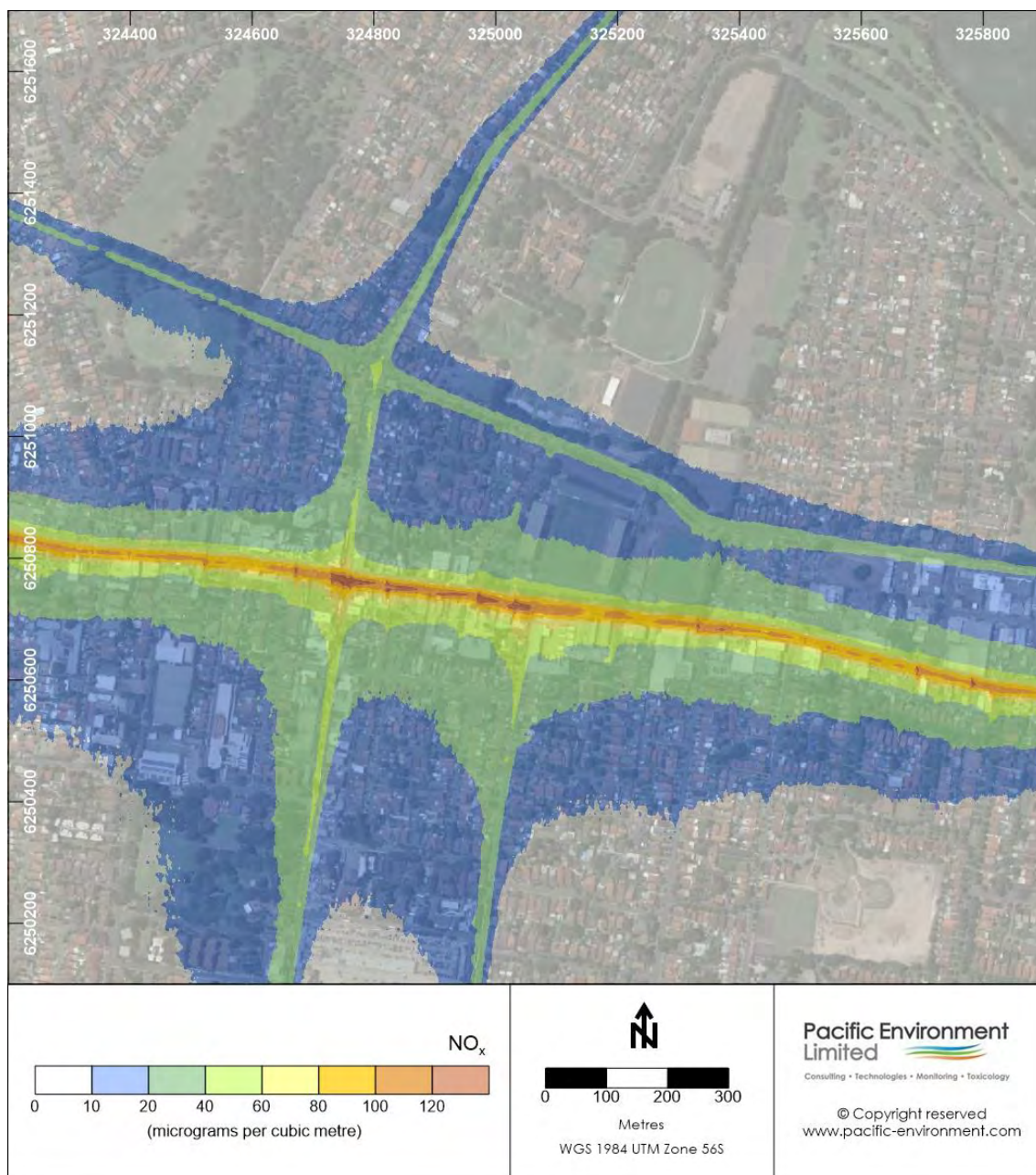


Figure F-17: Contour plot: average NO_x concentration (test GL-06)

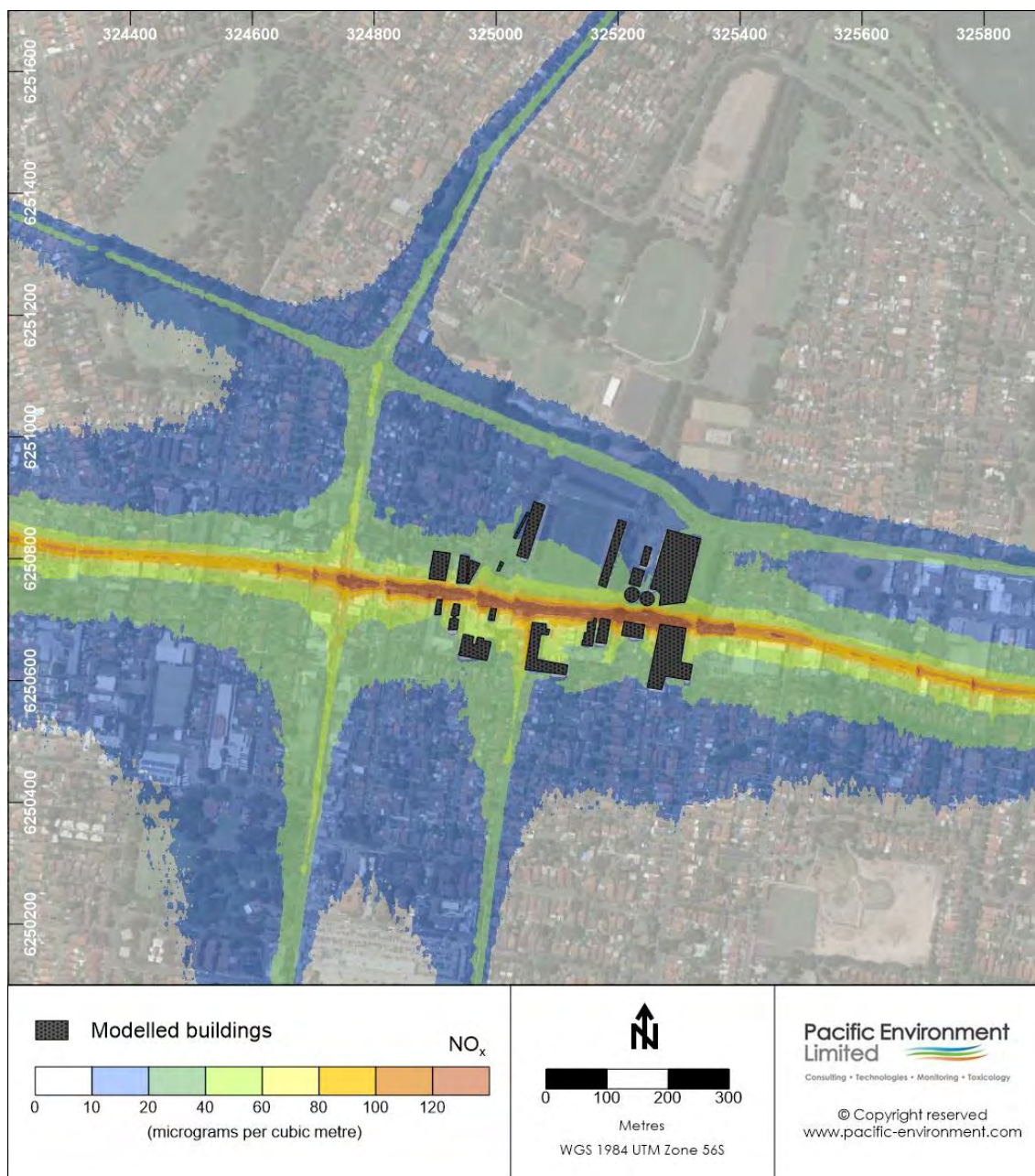


Figure F-18: Contour plot: average NO_x concentration (test GL-07)

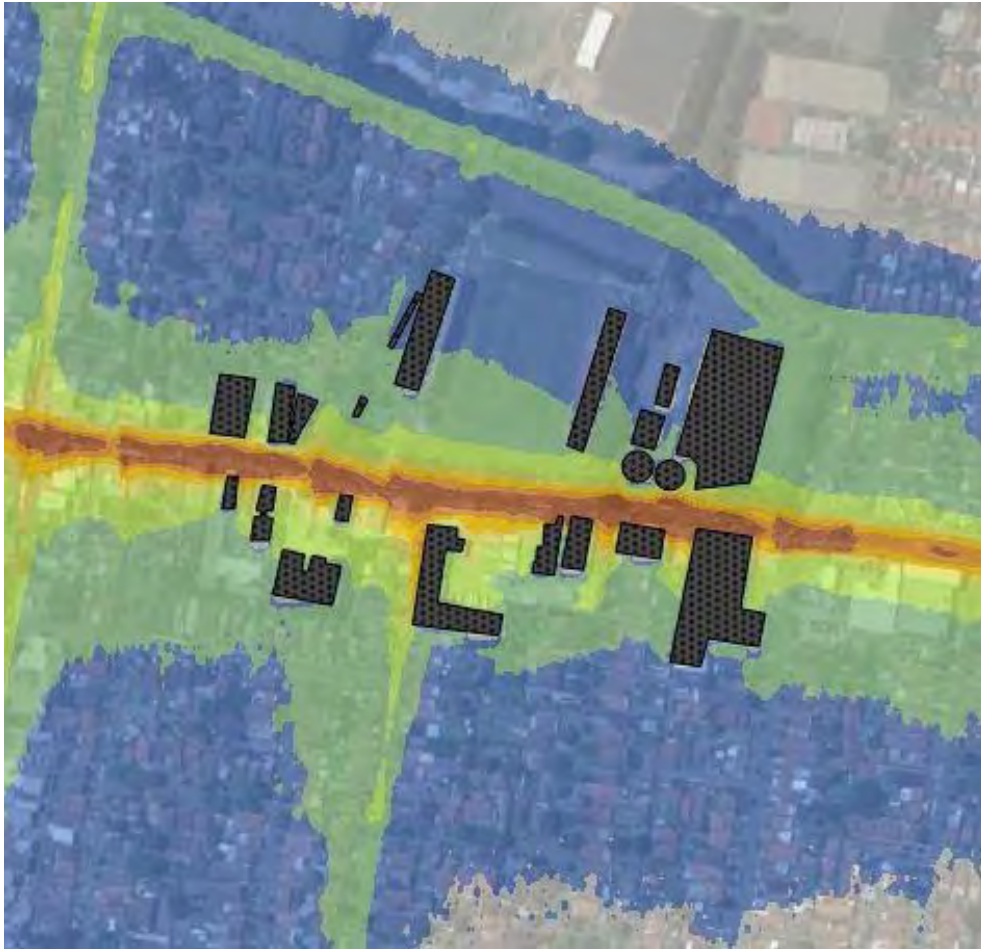


Figure F-19: Contour plot: average NO_x concentration (test GL-07, inset)

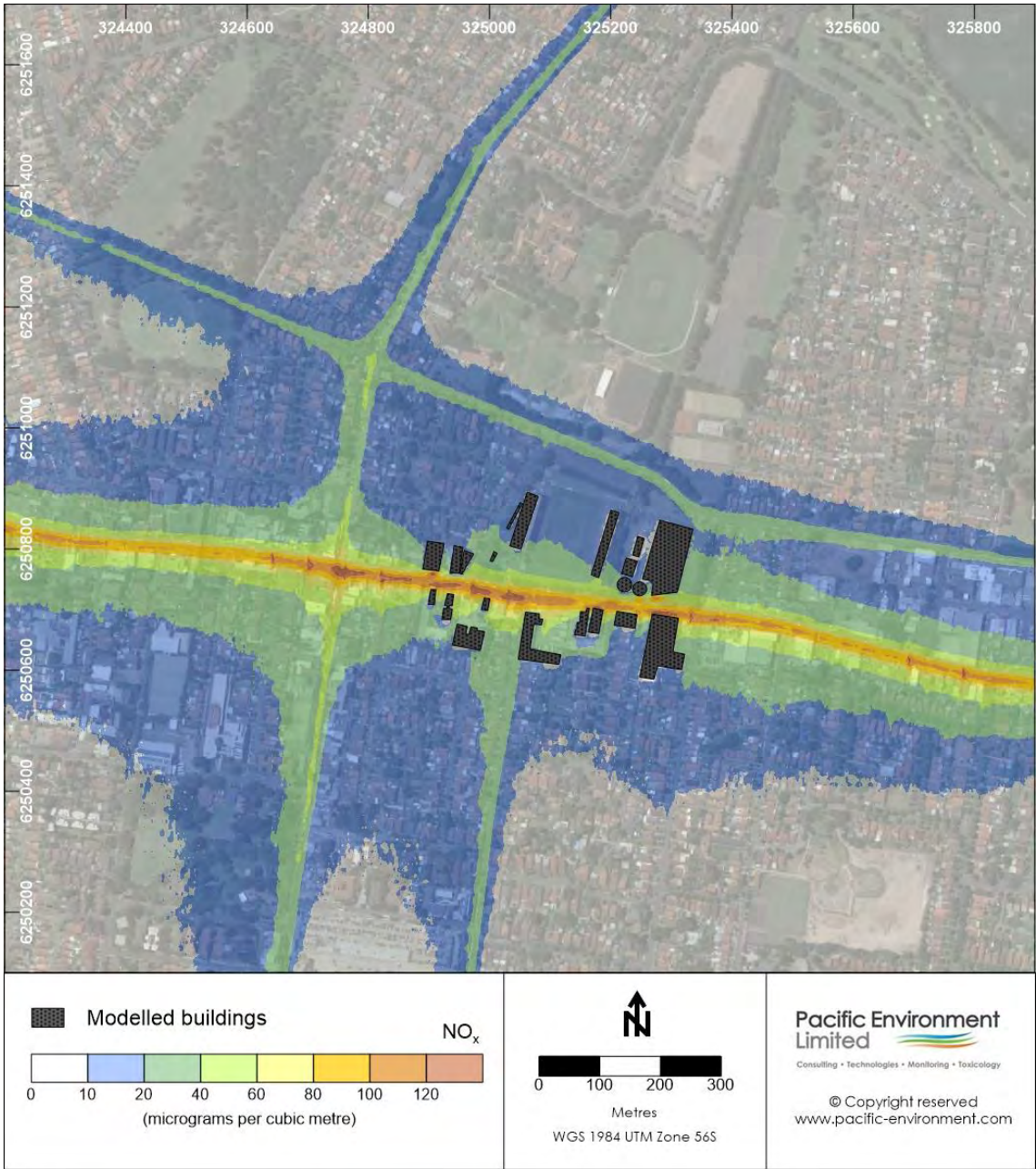


Figure F-20: Contour plot: average NO_x concentration (test GL-08)

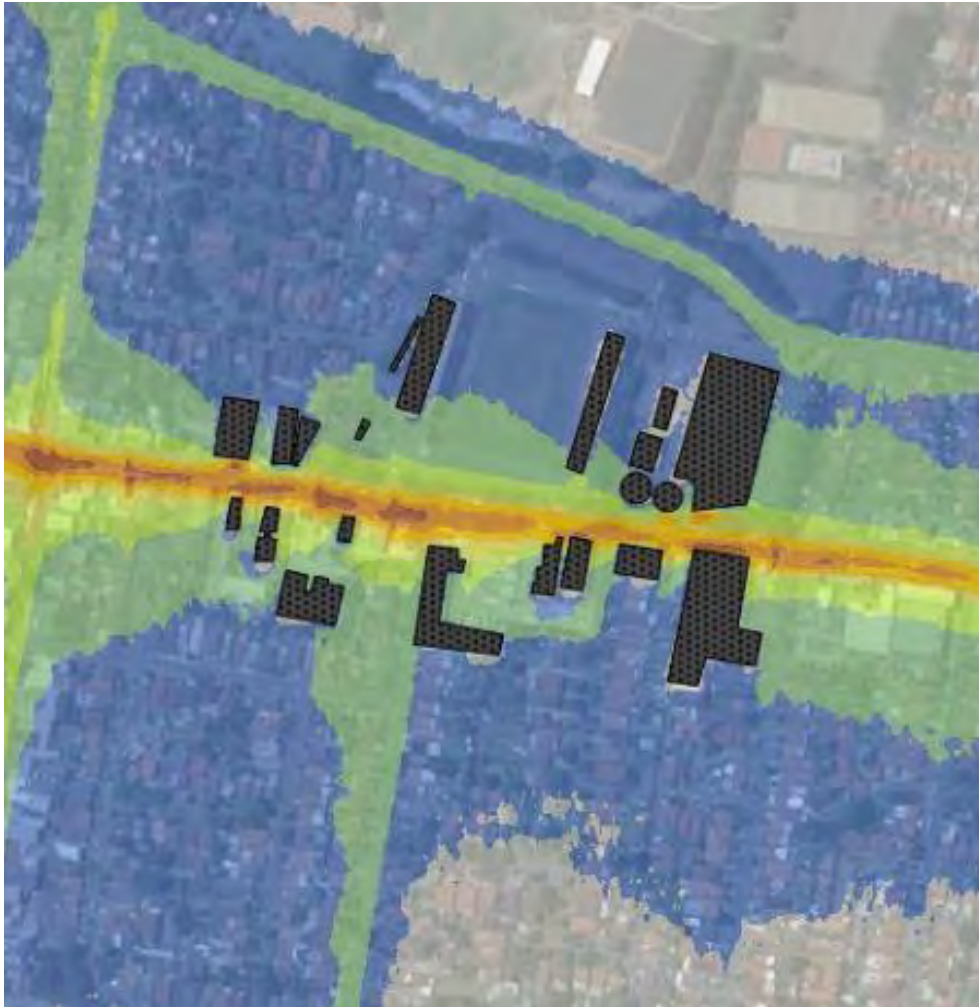


Figure F-21: Contour plot: average NO_x concentration (test GL-08, inset)

F2.1.1 Maximum 1-hour NO_x concentrations

Figures F-22 to F-32 show the contour plots for maximum 1-hour NO_x concentrations (model only, no background). For the tests including buildings (GL-07 and GL-08), additional plots are included to provide detail around Concord Oval.

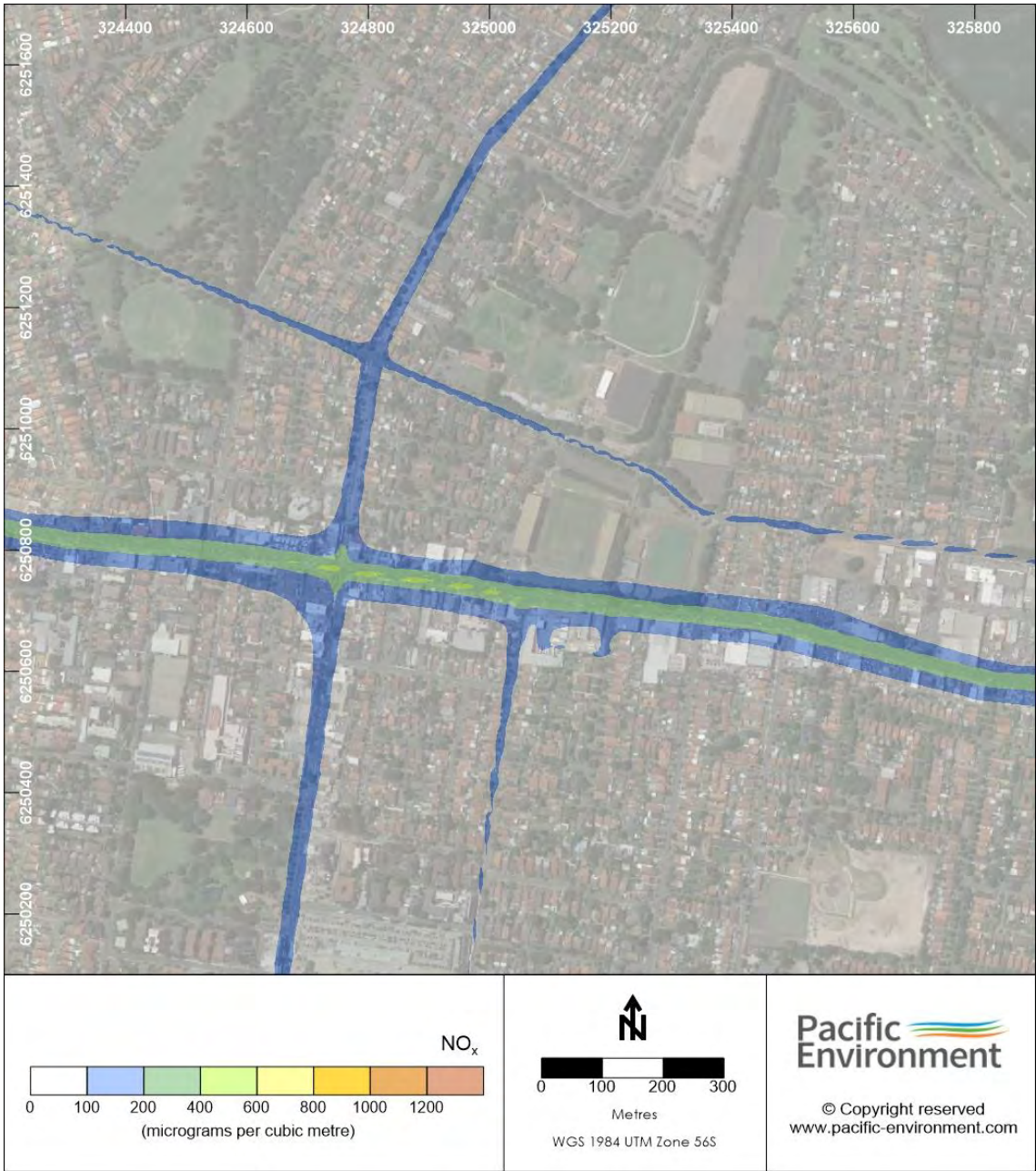


Figure F-22: Contour plot: maximum 1-hour NO_x (test C3-01)

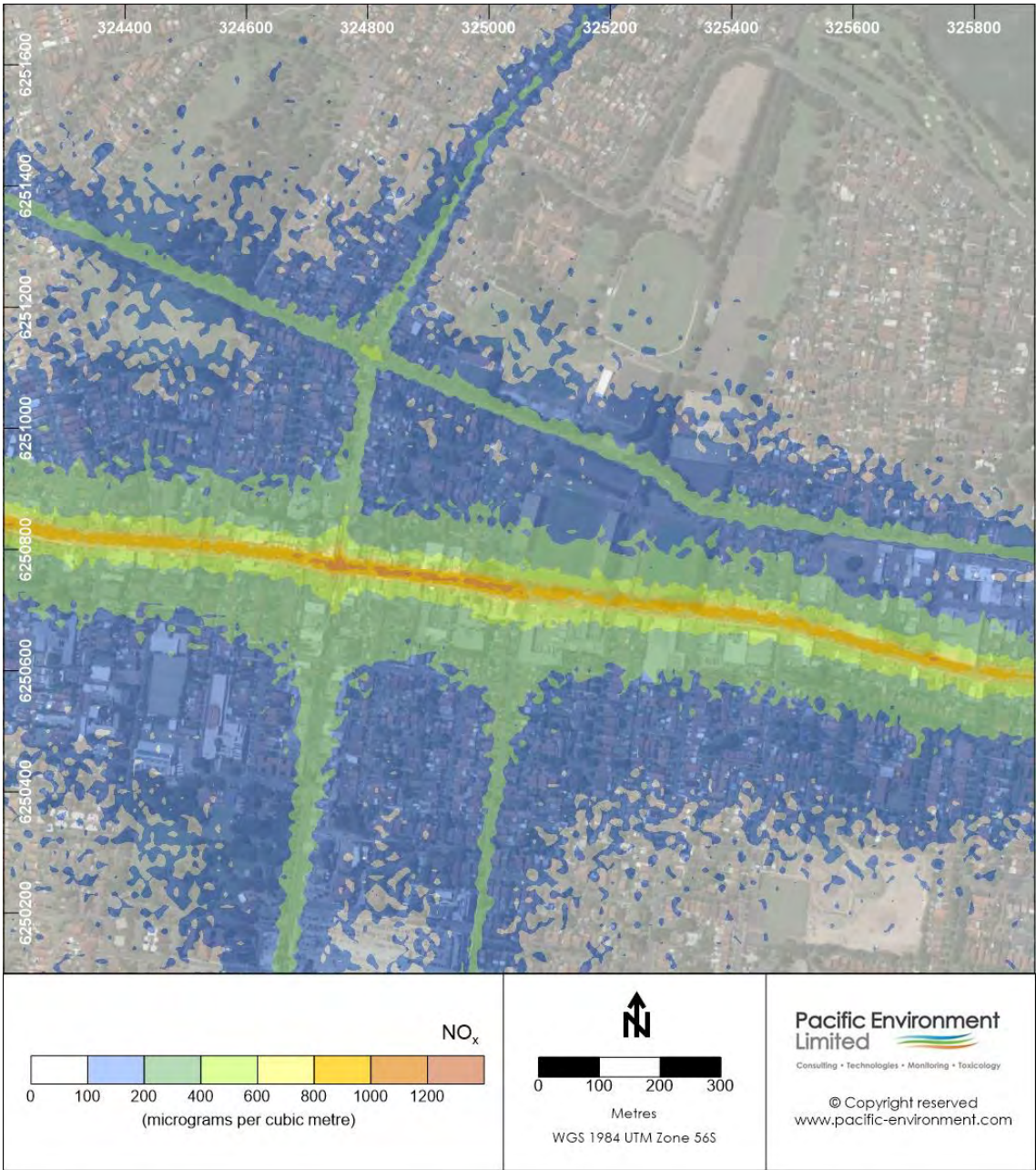


Figure F-23: Contour plot: maximum 1-hour NO_x (test GL-01)

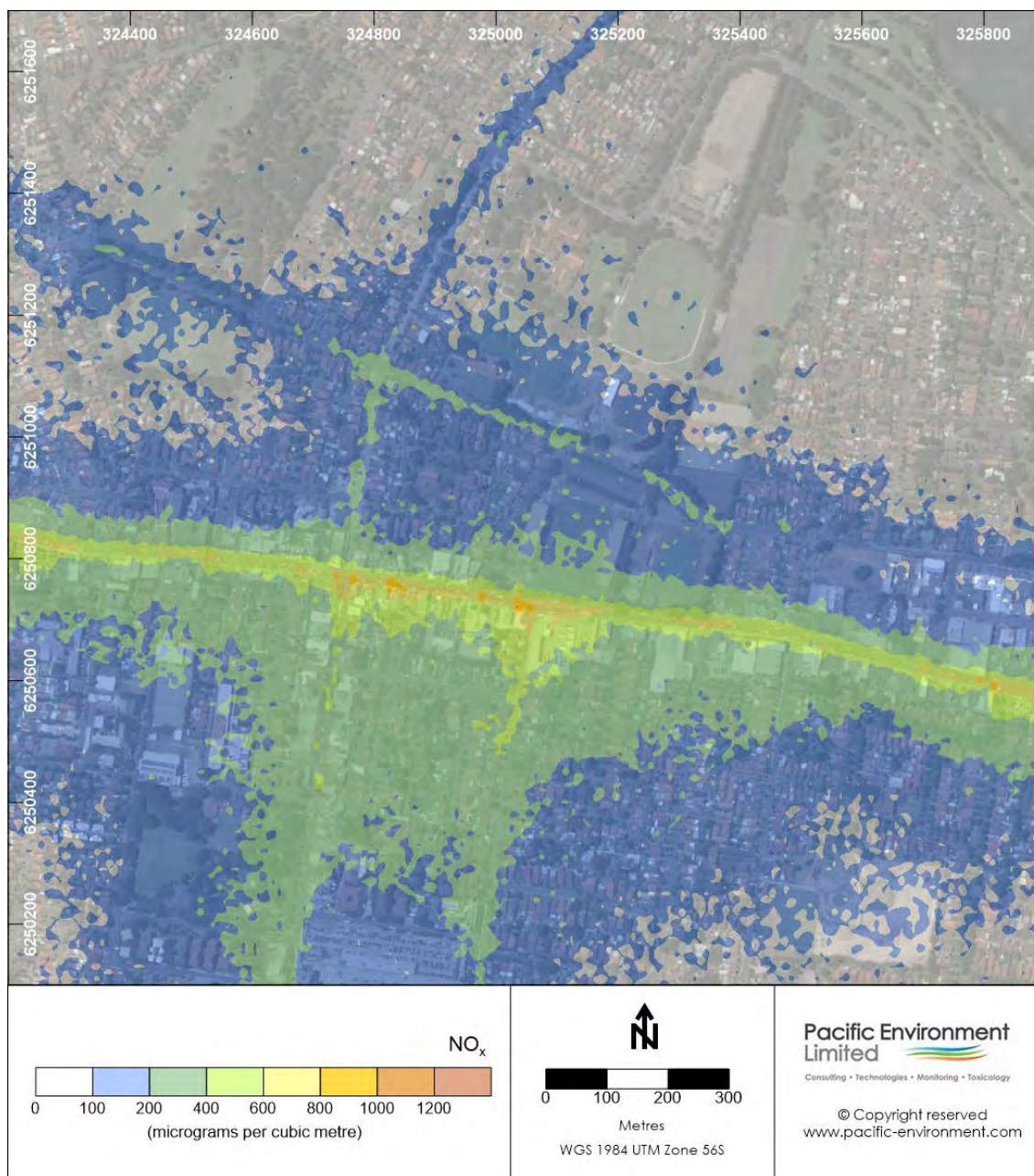


Figure F-24: Contour plot: maximum 1-hour NO_x (test GL-02)

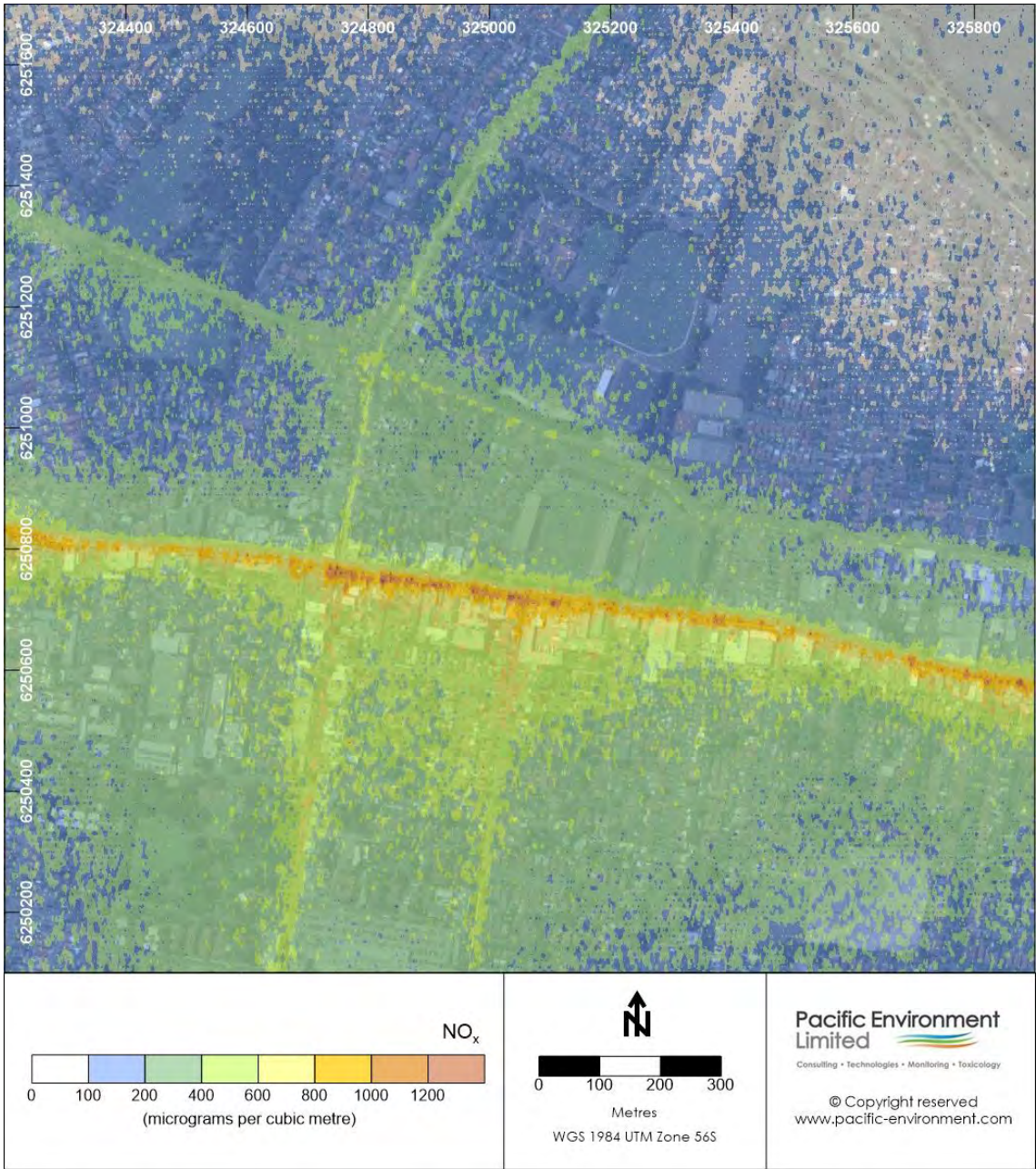


Figure F-25: Contour plot: maximum 1-hour NO_x (test GL-03)

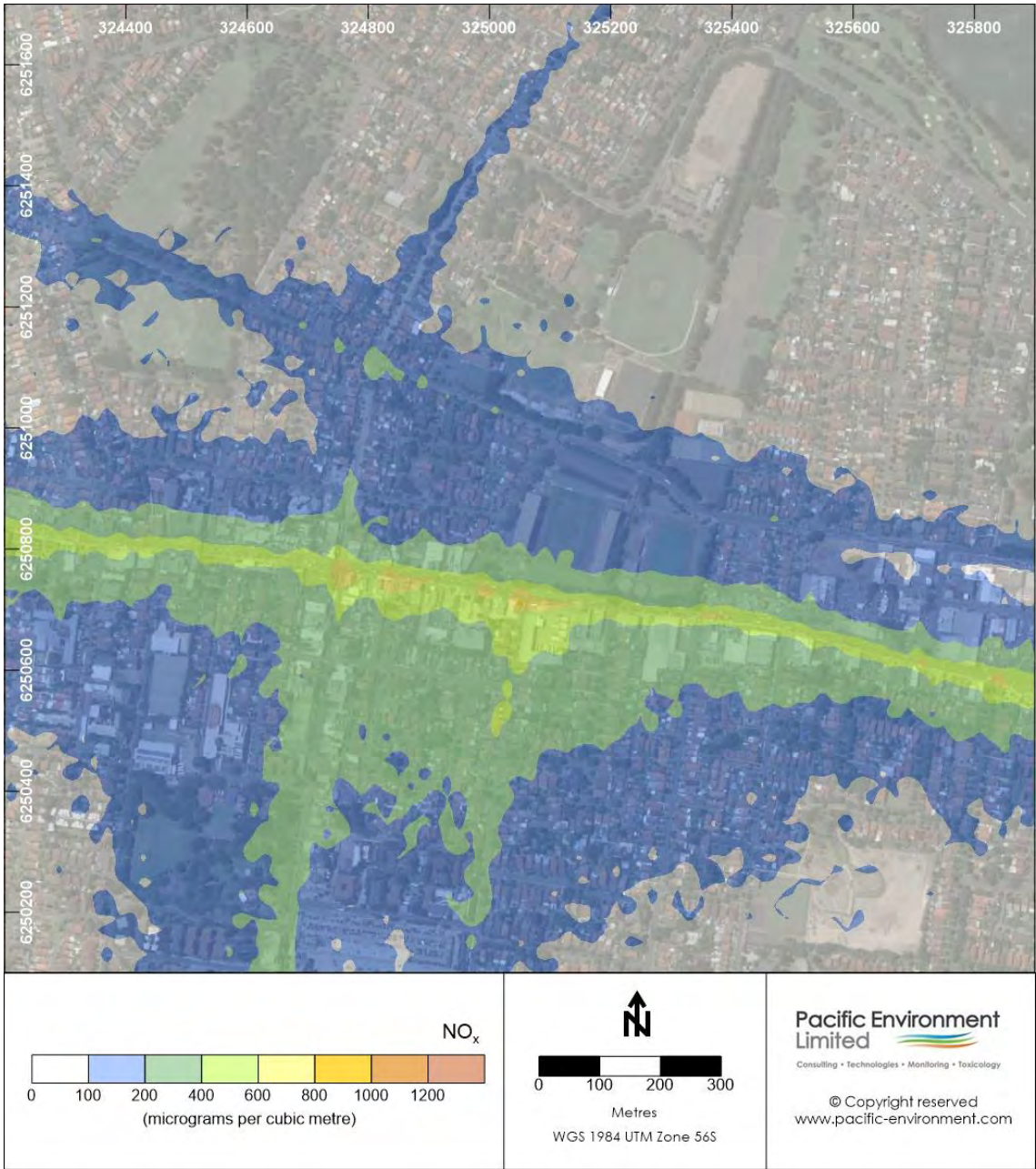


Figure F-26: Contour plot: maximum 1-hour NO_x (test GL-04)

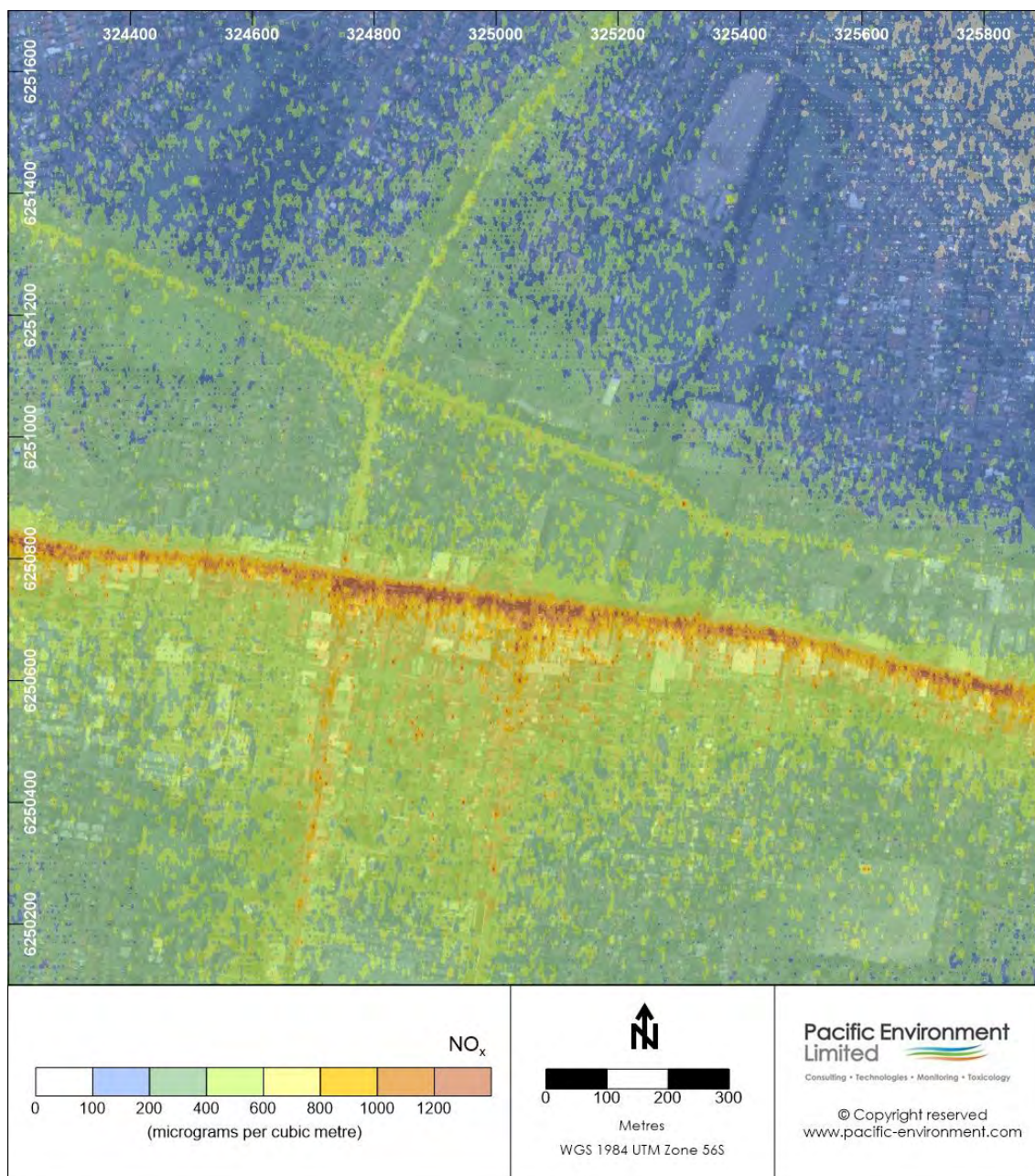


Figure F-27: Contour plot: maximum 1-hour NO_x (test GL-05)

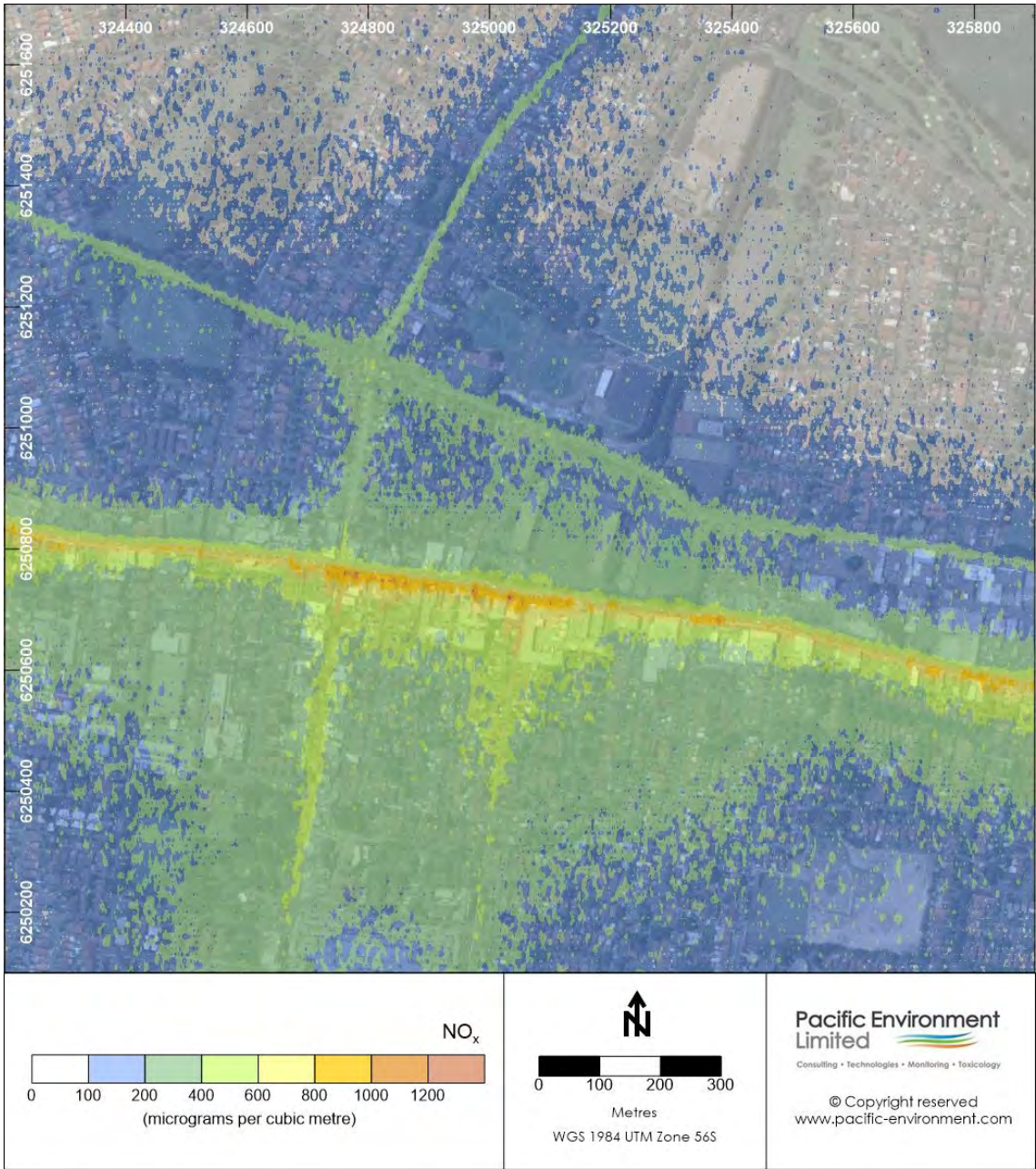


Figure F-28: Contour plot: maximum 1-hour NO_x (test GL-06)

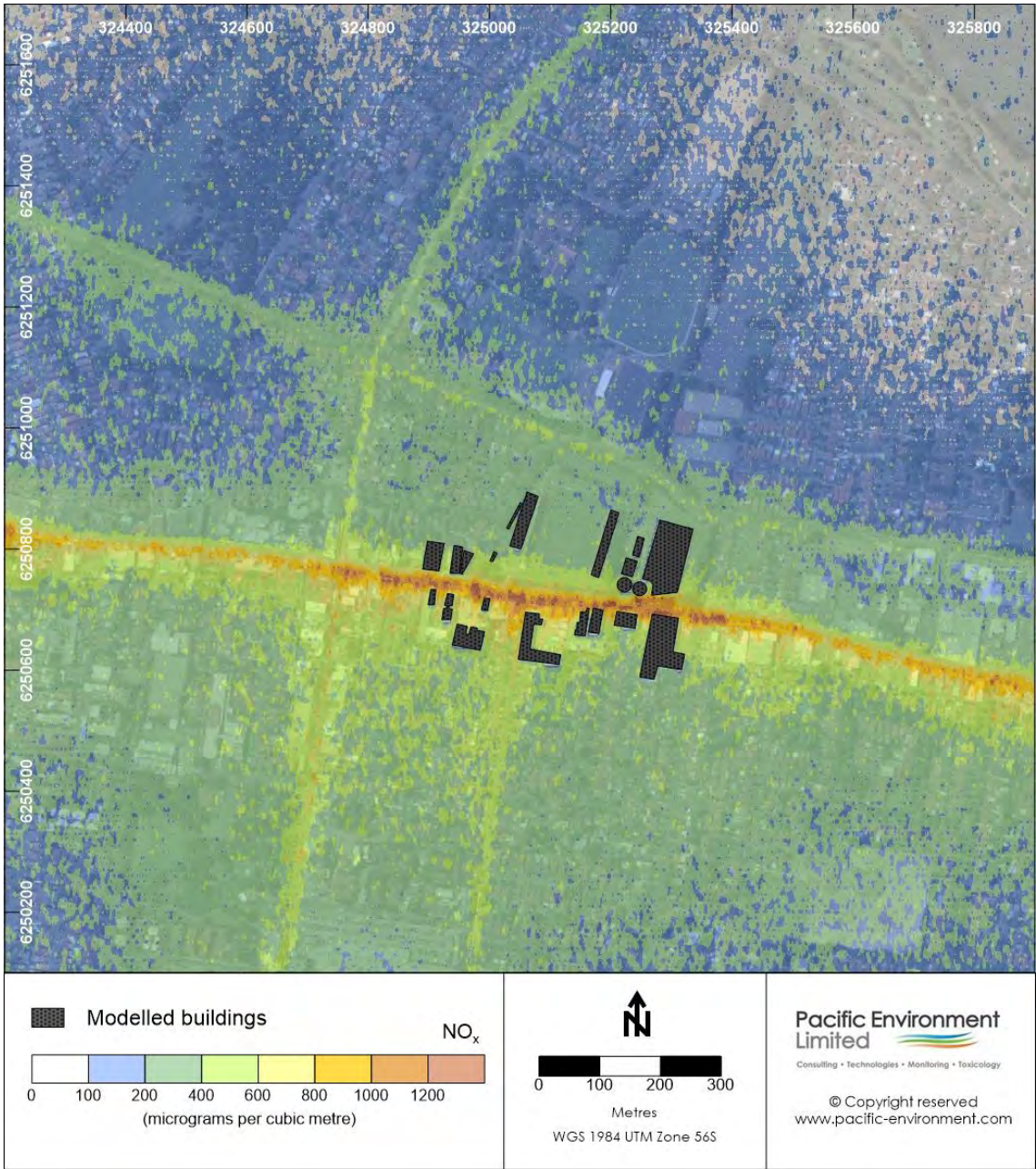


Figure F-29: Contour plot: maximum 1-hour NO_x (test GL-07)

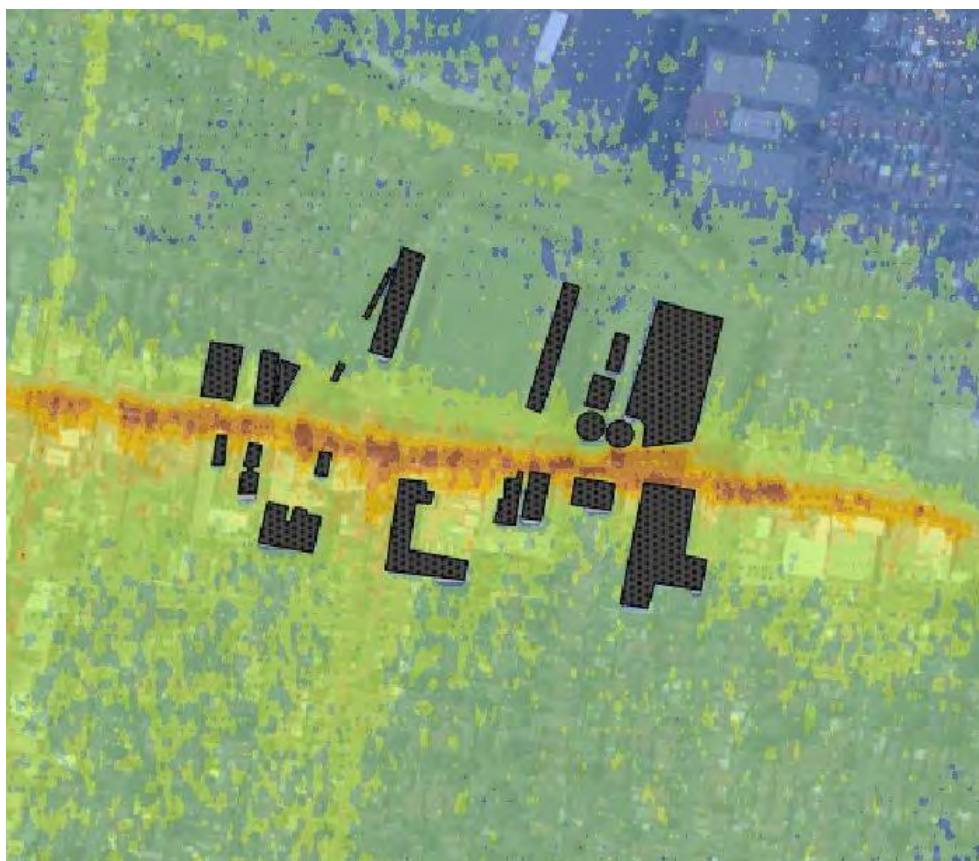


Figure F-30: Contour plot: maximum 1-hour NO_x (test GL-07, inset)

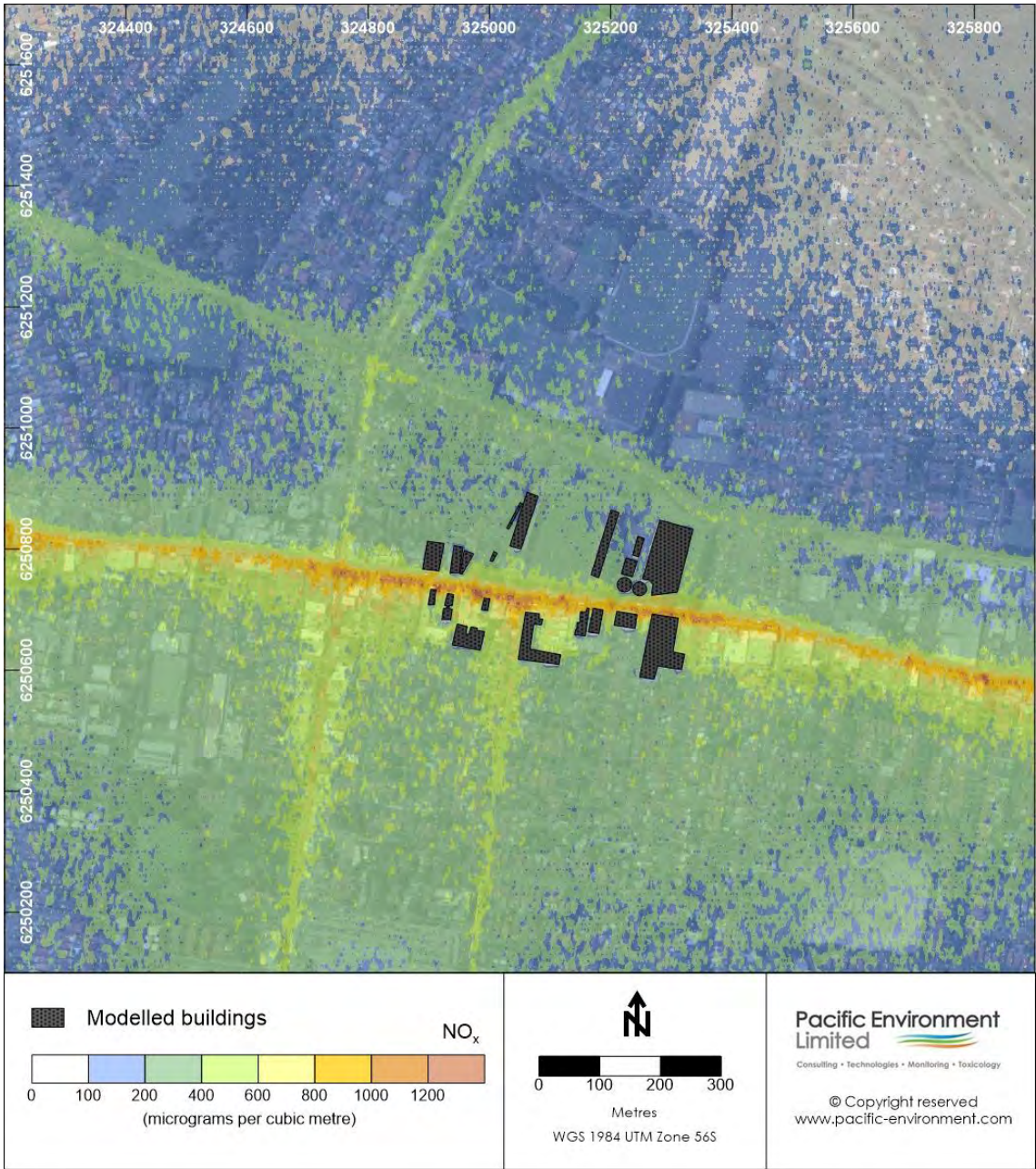


Figure F-31: Contour plot: maximum 1-hour NO_x (test GL-08)

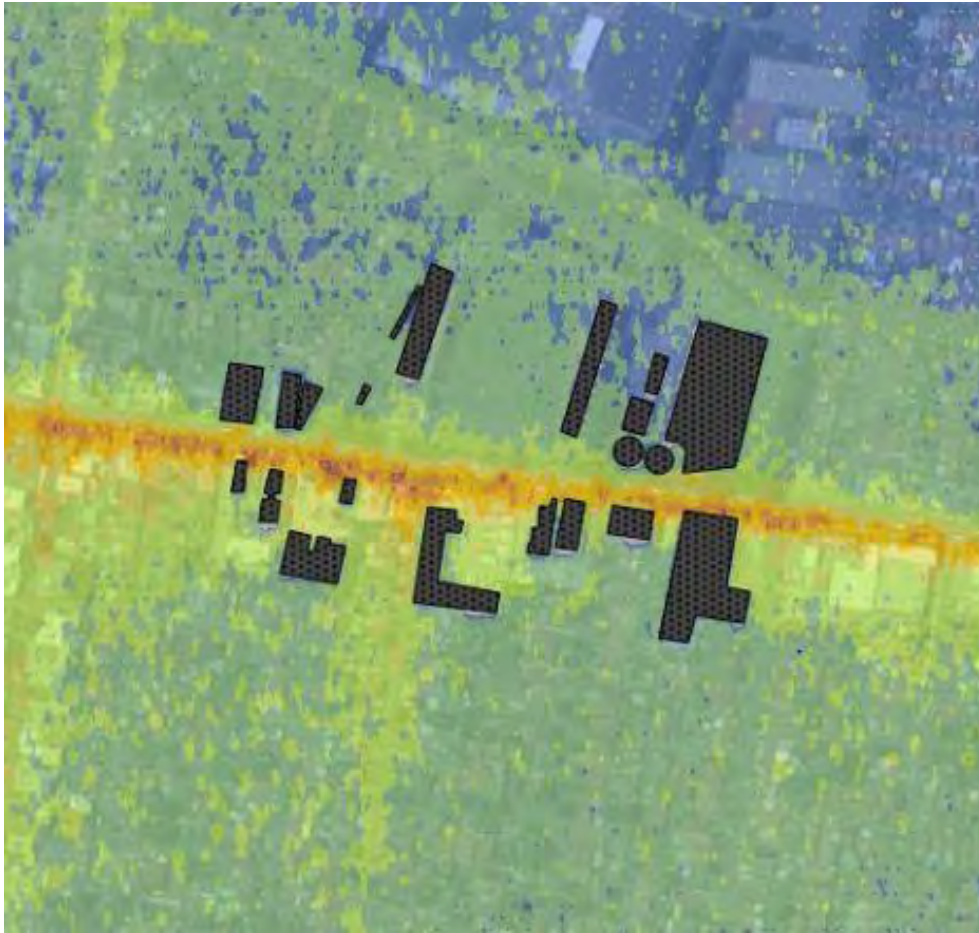


Figure F-32: Contour plot: maximum 1-hour NO_x (test GL-08, inset)