Not all speeds are created equal. Investigating the predictability of statistically downscaled historical land surface winds over central Canada.

by

Aaron Magelius Riis Culver
B.A.Sc., Queen's University, 2004

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the School of Earth and Ocean Sciences

Not all speeds are created equal. Investigating the predictability of statistically downscaled historical land surface winds over central Canada.

by

Aaron Magelius Riis Culver
B.A.Sc., Queen's University, 2004

Supervisory Committee

Dr. Adam H. Monahan, Supervisor
(School of Earth and Ocean Sciences)

Dr. Andrew J. Weaver, Departmental Member
(School of Earth and Ocean Sciences)

Dr. John C. Fyfe, Departmental Member
(School of Earth and Ocean Sciences)

Supervisory Committee

Dr. Adam H. Monahan, Supervisor
(School of Earth and Ocean Sciences)

Dr. Andrew J. Weaver, Departmental Member
(School of Earth and Ocean Sciences)

Dr. John C. Fyfe, Departmental Member
(School of Earth and Ocean Sciences)

ABSTRACT

A statistical downscaling approach based on multiple linear-regression is used to investigate the predictability of land surface winds over the Canadian prairies and Ontario. This study's model downscales mid-tropospheric predictors (wind components and speed, temperature, and geopotential height) from reanalysis products to predict historical wind observations at thirty-one airport-based weather surface stations in Canada. The model's performance is assessed as a function of: season; geographic location; averaging timescale of the wind statistics; and wind regime, as defined by how variable the vector wind is relative to its mean amplitude.

Despite large differences in predictability characteristics between sites, several systematic results are observed. Consistent with recent studies, a strong anisotropy of predictability for vector quantities is observed, while some components are generally well predicted, others have no predictability. The predictability of mean quantities is greater on shorter averaging timescales. In general, the predictability of the surface wind speeds over the Canadian prairies and Ontario is poor; as is the predictability of sub-averaging timescale variability.

These results and the relative predictability of vector and scalar wind quantities are interpreted with theoretically- and empirically-derived wind speed sensitivities to the resolved and unresolved variability in the vector winds. At most sites, and on

longer averaging timescales, the scalar wind quantities are found to be highly sensitive to unresolved variability in the vector winds. These results demonstrate limitations to the statistical downscaling of wind speed and suggest that deterministic models which resolve the short-timescale variability may be necessary for successful predictions.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank:

Dr. Adam H. Monahan, for being an ideal mentor and supervisor; his tremendous insight, support, patience, and knowledge have been invaluable.

My committee members, for sharing their time, suggestions, and knowledge.

Amy Torsney, for her love, support, and patience.

my family, for helping me reach this point in my life.

my colleagues in the Climate Modeling Lab, for their friendship.

the NSERC Collaborative Research and Training Experience Program, for funding my research.

Chapter 1

Introduction

On a global scale, surface winds are unceasing means of transport for momentum, heat, moisture, aerosols, and other atmospheric constituents. The transport takes the form of horizontal advection by the mean winds, as well as vertical and horizontal fluxes from turbulence and mixing that are heavily influenced by the winds. Accordingly, modeling the dynamics and probability distributions of the surface and boundary layer winds has emerged as a growing field.

Accurately modeling and characterizing surface winds is relevant for a number of practical problems. Modeling of microscale and mesoscale surface flow has long been a primary tool for pollution dispersion models. Accurate characterisation of the surface wind distribution is required for modeling loads on architectural designs. Microscale knowledge of the local wind resource is a leading factor in wind farm design and location selection. Furthermore, the strength of surface fluxes and momentum transfer has a direct dependency on wind speed. It follows that having a strong understanding of its dynamics, an accurate characterisation of its distribution, as well as robust means of predicting its development will have far-reaching benefits.

In recent years the economic value of the wind resource has garnered strong attention. On the heels of the Intergovernmental Panel for Climate Change's increasingly confident and pressing warnings on the widespread effects of anthropogenic carbon dioxide on the climate system, numerous jurisdictions have introduced measures to stimulate the decarbonisation of society's energy supply. Over the last two decades

a large industry has grown in Europe, and more recently in North America, whose product is the conversion of wind energy into electricity. In Canada, provincial governments have introduced policy initiatives to increase power generation via renewable sources, particularly wind and solar farms (Ontario Power Authority, 2010; BC Hydro, 2011). Estimates of the power in the wind resource are calculated from probability distributions of the local observations. Consequently, it is the norm that feasibility studies, despite being forward-projecting analyses, rely on characterisations of the historical wind resource. Some qualification of the future wind resource in a changing climate would broaden the utility of these studies. Furthermore, the possibility of predicting variability of the wind resource on a range of timescales offers a significant advantage for maintaining sufficient cash flow. This is one application where successfully predicting the surface winds has tremendous potential.

Weather and climate predictions are ultimately produced by dynamical models of the global-scale atmospheric circulation, i.e. General Circulation Models. Dynamic modeling is inherently complex and computationally intensive. At the heart of such a model, the equations governing the conservation of energy, momentum, and mass are solved numerically (e.g. Holton, 2004). The result is a four-dimensional space-time grid of meteorological variables. The resolution of current dynamic models is widely varied, as are the required computational resources. No matter what the resolution, as a result of the discretized representation of the state of the atmosphere, there will always be dynamic interactions and topography at sub-grid scales that are parametrized. This is particularly the case with models that simulate flow on continental and global scales. These models are particularly well-suited for accurately describing the geostrophic flows aloft, where dynamical spatial scales are relatively large, but their performance at the surface and within the planetary boundary layer will be limited by the approximations made necessary by their discretisation (Giorgi and Mearns, 1991).

There are numerous means of providing high resolution, local predictions of the climate. These can be grouped into two broad but distinct categories: dynamic downscaling and statistical downscaling. Dynamic downscaling is a means of circumventing the errors introduced by parametrizations in large-scale climate models by locally increasing the spatial and temporal resolution. Computing limitations result in the increased resolution coming at the cost of a limited spatial coverage. Typically

the upper limit of the scale of regional downscaling models is on the order of North America's size. These models do demonstrate improved accuracy within the planetary boundary layer, but there are basic features of the land surface winds that are poorly resolved (He et al., 2010). Furthermore, areas of strong topographic influences may still not be accurately resolved in regional models. This issue may be addressed by running a finely resolved local model, but it may be impractical to do so for many sites at once.

In this regard, statistical downscaling is a natural complement to dynamic modeling. Via statistical methods, local surface variables may be related to the free tropospheric circulation that is more accurately resolved by the dynamic models. Surface winds are determined by both the large-scale circulation and local features, such as topography and land/water contrast. Statistical downscaling (SD) methodology seeks to identify the information in the large-scale flow aloft that is relevant to the fluctuations of the local surface anomalies. With the relevant information in the circulation aloft identified (the predictors), SD then models the relationship between the statistics of the local variable (the predictand) and the statistics of the predictors.

The performance of SD, typically characterised by the percent of the anomaly's variance explained by the prediction, is directly related to the anomaly's sensitivity to the large-scale circulation. If the anomaly is predominantly a function of the large-scale circulation, it will be well predicted. If instead, the observations are subject to strong local influences, perhaps interactions with neighbouring structures or topography, the local wind is likely to be poorly predicted. In this case, it may be necessary to rely on microscale or mesoscale computational fluid dynamics models, which have the resolution to resolve the nearby structures to downscale the mesoscale flow and predict the time-varying anomaly.

Across southern Canada there are numerous observation stations whose records extend from the mid-20^{th} century to the present. The potential for successfully predicting surface winds at these locations using SD can be assessed by searching for similarities across sites that are separated by distances on the order of a thousand kilometers, as such similarities would reflect the common influence of large-scale flow aloft. For instance, despite the large spatial extent between the sites shown in Figure 1.1, these sites may be qualitatively grouped into two sets with distinct seasonal cycles. The monthly means observed in four sites representative of the two seasonal

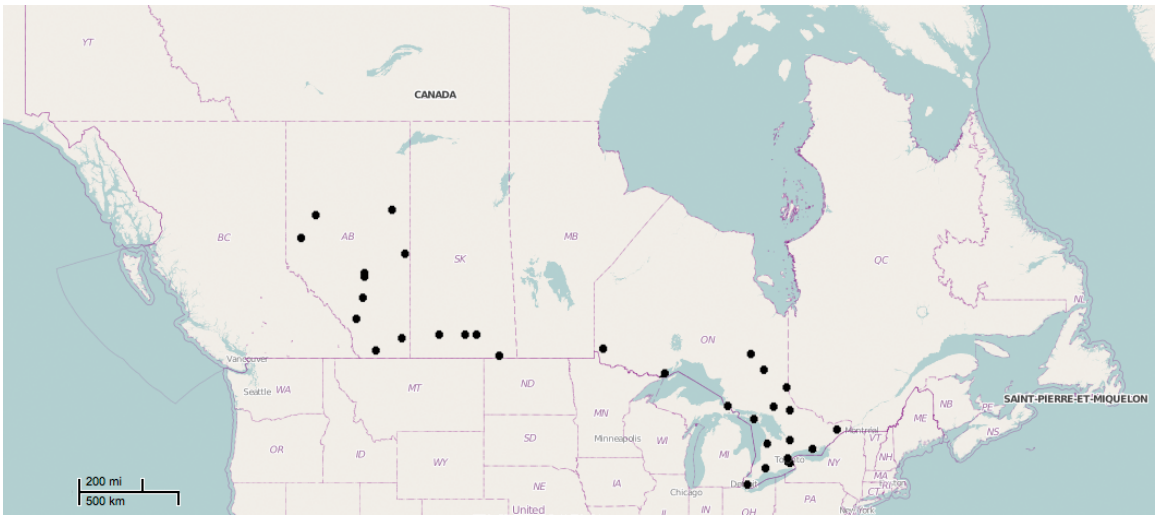

Figure 1.1: The locations of long-standing weather observation stations in Alberta, Saskatchewan, and Ontario.Map obtained from Braxmeier (2011)

cycle patterns are shown in Figure 1.2. The central and northern Alberta and Ontario sites have a relatively steady seasonal cycle, with slightly lower speeds in the equinoctal seasons. On the other hand, the southern sites have a pronounced annual cycle, with low mean wind speeds in the summer and high mean wind speeds in the winter months. The seasonal cycle is not the subject of this study, and recent work has shown that its statistical prediction is somewhat arbitrary (Curry et al., 2011). Nevertheless the similarities between sites across such expansive distances suggests a possible common connection to the annual cycle of the large-scale circulation aloft.

A more finely resolved investigation of a common connection to the circulations aloft would be to examine the spatial extent of the correlation of fluctuations on monthly timescales among the observations. To avoid resolving the shared seasonal cycle in the linear relationship between sites, the seasonal cycle may be removed from the time series, or the time series may be split into seasons. Adopting the latter procedure, in Figure 1.3 the correlations of wintertime (DJF) monthly mean wind speeds at Calgary AB, Moose Jaw SK, and Sault Ste. Marie ON and the corresponding monthly means at all other sites are displayed. It can clearly be seen that sites in Ontario (bottom panel) have a modest degree of correlation despite differences in local topographic features. Wind variability at Sault Ste. Marie ON, which neighbours Lake Superior and Lake Huron, correlates well ($\rho > 0.45$) with sites located deep within the northern forests of Ontario as well as sites in the southern prairies of Saskatchewan.

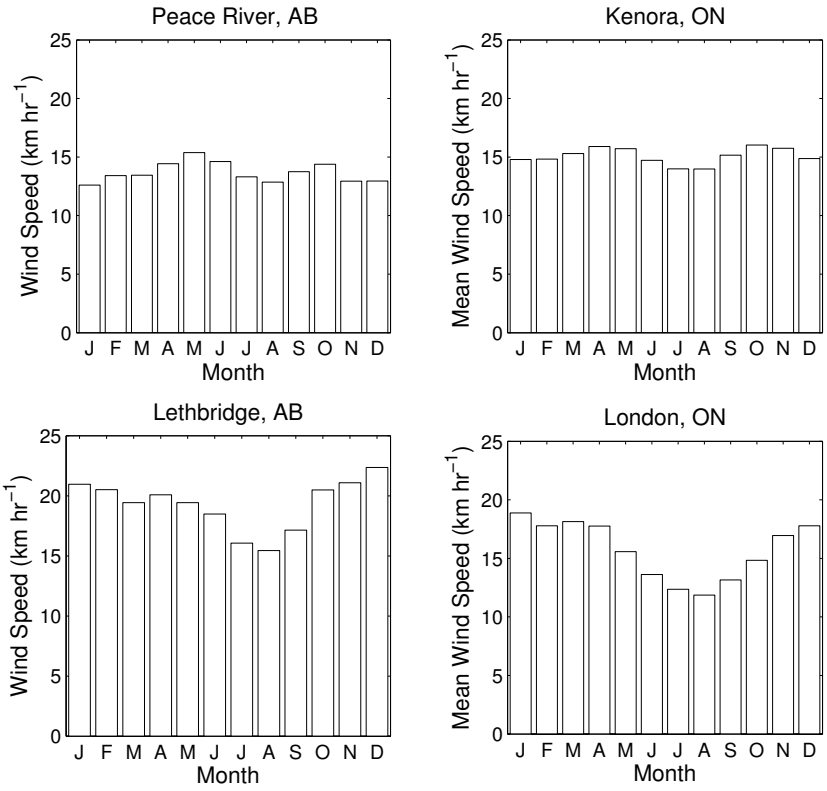

Figure 1.2: The monthly mean wind speed [km hr^{-1}] of four locations representative of the two distinct seasonal cycles observed at the sites in Figure 1.1.

The fluctuations in the wintertime monthly mean wind speeds of Calgary AB are in large part shared with the winds observed in the northwestern prairies of Alberta, the southern prairies of Alberta and Saskatchewan, and Ontario. Correlations do not generally scale with the distance between sites. This is particularly evident in Ontario, where higher correlations were found with remote sites, relative to nearby ones. The large spatial extent of the correlations observed provides evidence that these sites are sufficiently influenced by the large-scale circulation to warrant the use of statistical downscaling.

St. George and Wolfe (2009) provided a first quantitative estimate of the potential predictability of wind speeds at the southern prairie sites shown in Figure 1.1. Their study assessed the linear relationship between the regional winter wind speeds in the southern Canadian prairies (SCP) and variability of the El-Niño Southern Oscillation (ENSO) 3.4 index (N3.4), a measure of the sea-surface temperature anomalies in the eastern half of the equatorial Pacific Ocean. ENSO is known to have important effects

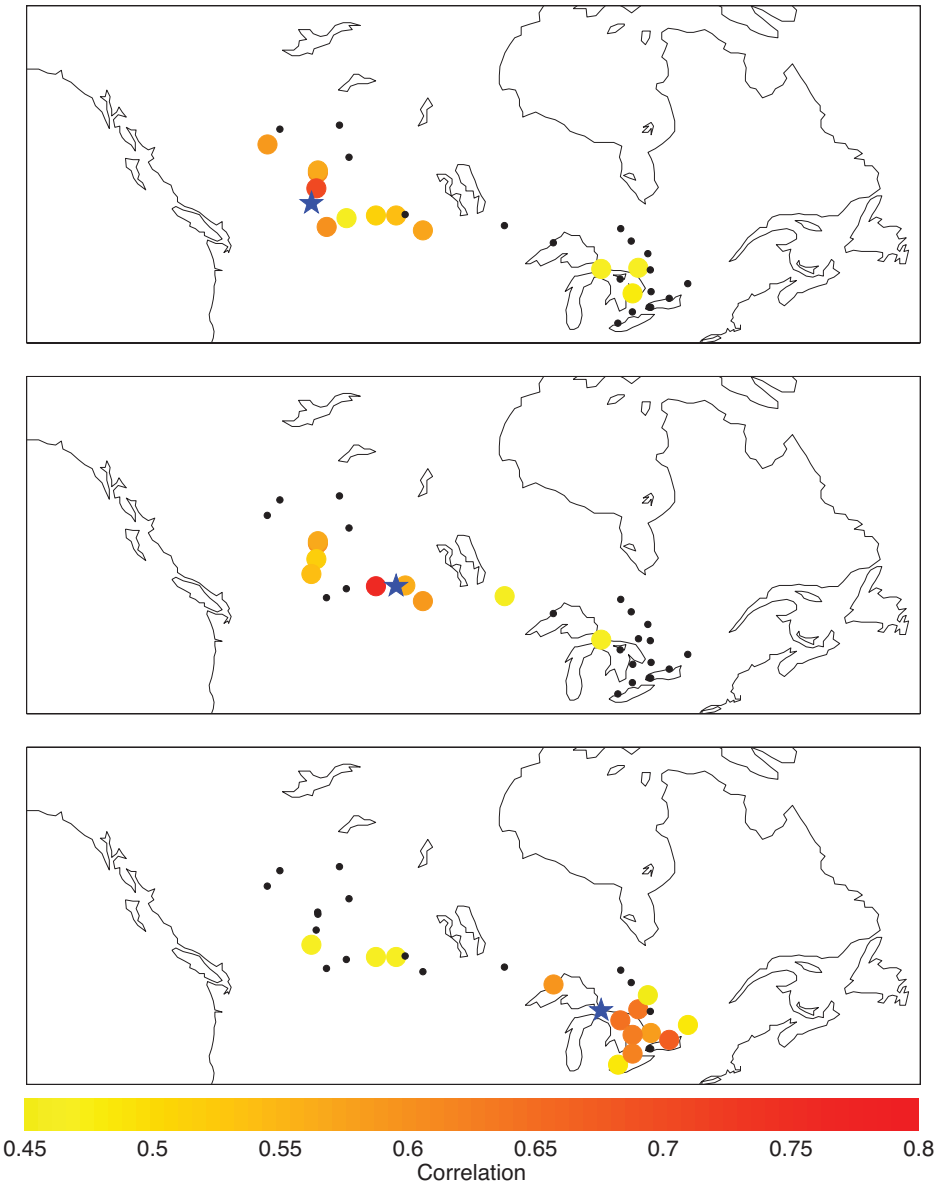

Figure 1.3: The correlations of all sites' wintertime monthly mean wind speeds with those of three specific sites: Calgary AB (top), Moose Jaw SK (middle), and Sault Ste. Marie ON (bottom). The blue star indicates the location of each of the specific sites. Only correlations greater than 0.45 are shown.

on tropospheric circulations over North America (Allan et al., 1996). St. George and Wolfe (2009) found that the regional mean winter winds over the SCP exhibited a "significant and inverse" relationship with the sea-surface temperatures in the Niño 3.4 region. While the relationship was certainly statistically significant (at the $p = 2 \times 10^{-5}$ level using a one-tailed t-test), it was not particularly strong. The variance in the sea-surface temperature was found to describe only 25% of the variance in the regional SCP wind speed. Furthermore, this estimate of predictability was not based on a cross-validated prediction, but only an examination of the linear statistical relationship between the two variables. The result can therefore be interpreted as an upper estimate of the true prediction skill attainable when predicting regional SCP winds with the N3.4 index. The correlation of SCP winds with a large-scale perturbation of the tropospheric flow also provides further evidence that at least a modicum of success should be expected if SD is applied to the surface winds over this region.

The techniques used for statistical downscaling are diverse. As a first classification, such methods may be grouped into two distinct types: linear techniques and nonlinear techniques. Within each of these categories are statistical methods of varying complexity. Both classes of techniques have been applied to the downscaling of surface winds. Some studies have suggested that these two types of models compare quite well, with no real added advantage from the increased complexity of nonlinear models (Zorita and von Storch, 1999). However, this result might be an overgeneralization. Davy et al. (2010) applied a random forest algorithm (a linear statistical downscaling method that formed predictions from an ensemble of regression trees) to the prediction of wintertime and summertime surface wind variability at an Australian site. In this instance, the algorithm was found to outperform multivariate linear regression in both seasons (but only marginally in the latter). To further confuse the assessment of the relative merits of different SD techniques, nonlinear regression was found to outperform linear regression in the prediction of temporal variability of daily wind speeds in France (Najac et al., 2009; Salameh et al., 2009). These studies suggest that there are instances where linear methods will outperform non-linear methods and vice versa; as well as instances where more complex models have outperformed simpler ones. There is no evident systematic reason as to the circumstances in which one technique should be superior to another.

With any of the afore-mentioned SD techniques, an increase in model complexity (and hence number of statistical model parameters) allows for more complex relationships to be modelled. With increased complexity, statistical models require data in larger amounts and of higher quality to robustly estimate their parameters. Evidence of this is seen in the Curry et al. (2011) critique of the probabilistic downscaling approach taken by Pryor et al. (2005). In lieu of predicting interannual variability (a dataset with a relatively high number of statistical degrees of freedom), Pryor et al. predicted the seasonal cycle of the Weibull-distribution parameters describing the monthly wind speeds (with a small number of statistical degrees of freedom). Curry et al. (2011) applied a similar approach to wind observations across British Columbia and noted a strong lack of robustness to the statistical model of Pryor et al.. In particular, spurious predictions of the seasonal cycle of winds were found to be possible with any sort of seasonally cyclic variability, without the existence of a causal connection. The lack of a casual connection would not (technically) be an issue if the present climate persists, but is an essential problem for the use of statistical downscaling in assessing climate change. The results of Curry et al. (2011) further justify the focus on predictions of interannual and interseasonal quantities.

This study will make use of a simple linear statistical downscaling technique: multivariate linear regression, to assess the predictability of interannual and interseasonal surface winds over three Canadian provinces. The statistical model is neither novel nor complicated, but due to its simplicity it is relatively robust. Past statistical downscaling studies have found vector winds to have better predictability than wind speeds, but until recently (e.g. Monahan, 2012b) no discussion has been given to why there should be such a difference. In addition, the current literature provides little consensus on which averaging timescales will garner better predictability. Past work predicting surface winds in highly topographically influenced regions of France found weekly timescales to be better predicted than daily or hourly timescales (Salameh et al., 2009). In contrast, recent studies of the sea surface winds off the coast of British Columbiafound predictability to be greatest at daily timescales (Monahan, 2012a). Furthermore, the work by Salameh et al. (2009) suggested some projections of the wind vector are better predicted than others because of topographic influences constraining the dominant wind direction. A similar result was seen in van der Kamp et al. (2011), where maximum prediction skill of the vector projections typically aligned with topographic features such as ocean straits and mountain ranges.

Each of these traits in surface wind predictability will be examined in the context of the Canadian sites considered.

This study considers the prediction of historical surface winds over 31 observational sites, spanning topographically-diverse regions of central Canada. In doing so, the study accomplishes the following objectives:

1. to quantify the statistical predictability of land surface winds across Alberta, Saskatchewan, and Ontario,
2. to assess the sensitivity of the prediction skill to;
    - vector versus scalar predictands (e.g. wind components versus wind speed),
    - the four calendar seasons,
    - the statistical (averaging) timescales,
    - the location of the predictands, and lastly,
    - the wind regime, as defined by how variable the vector wind is relative to its mean amplitude.
3. to provide an interpretation of these results using an idealized probabilistic model.

Predictions are made of historical winds only; accordingly the predictability results and discussion that follow may be accurately thought of as exploration of the linear relationship between the mid-tropospheric means and the statistics of the surface winds.

In this study, the predictors are the dominant modes of variability of several tropospheric variables. The predictands are the means and standard deviations of the historical scalar and vector surface winds. The goal of this study is not to compare predictive skills among different downscaling approaches, nor make predictions of future winds. Instead, this study will focus on linear predictability and its variability across spatial location and timescale. The following chapter details the methodology and data used in this study; specifically the climate data used, the selection of the predictor variables, the construction of the predictors, and the validation of the statistical model. The third chapter presents the prediction skill realized with the

statistical model. The predictability of the surface winds are discussed in the following chapter. Conclusions and a discussion of the results in the context of the SD literature are presented in Chapter 5. Typically the surface wind speed and vector components will be referred to as the scalar wind and vector wind respectively. A glossary of frequently used terms and symbols used through the current study are included in Appendix A.

Throughout the study, the notation $\overline{y}$ is used to denote the mean of a quantity y on a particular timescale, and σ_y is used to denote the associated standard deviation of y on sub-averaging timescales. That is, if $\overline{y}$ represents a monthly mean of y, σ_y represents the standard deviation of y within that month. The notation $mean(y)$ and $std(y)$ is used to denote long-term statistical features of y.

Chapter 2

Methodology and Model Development

2.1 Data

This study used a statistical downscaling approach (multivariate regression) to construct statistical relationships between historical observations of surface winds (the predictands) and the large-scale flow aloft (the predictors). The climate data used in this study consisted of in situ wind velocity observations from anemometers, a regional re-analysis product, and a global re-analysis product.

2.1.1 Predictands

Observations of surface winds from thirty-one stations were selected for the study, ten from Alberta, four from Saskatchewan and seventeen from Ontario (see Figure 2.1). All stations are located at airports and have at least 30 years of observations. Metadata for the weather stations used in this study can be found in Appendix B, including station location (latitude, longitude, elevation), data duration, instrument type, and documentation of anemometer height changes. The observations have varying degrees of missing values. Some are recently inactive, dropping off-line as of the early 2000s, others have dropped from 24 hours of operation to eight or sixteen hours

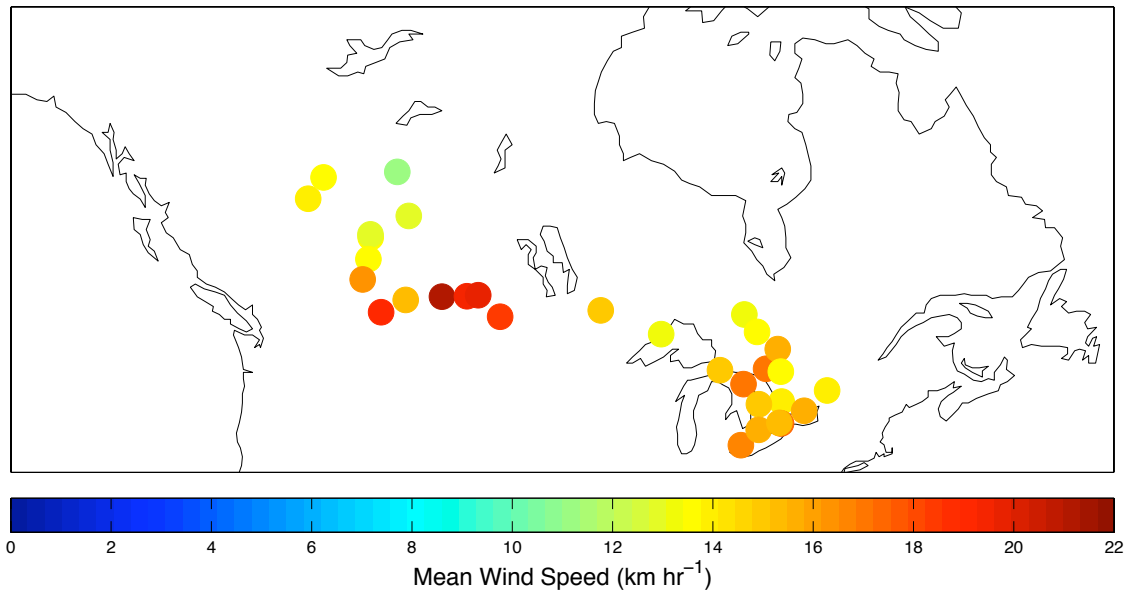

Figure 2.1: The locations and long term mean wind speeds of the observation records used in this study's statistical downscaling. The colour of the circle corresponds to the long term mean wind speed.

and are inactive during weekends. Provided their availability, surface observations dating from 1953 to 2006 were used to construct statistical relationship with the large-scale circulation.

A large majority of the anemometers at the stations shown in Figure 2.1 have been re-located to a standard height of 10 meters (at present, the anemometers at Fort McMurray AB, Earlton ON, North Bay ON, Muskoka ON, and Toronto Island ON remain at heights of 13.1 m, 19.5 m, 13 m, 26.9 m, and 11.4 m respectively). As a result, at some point in their data record, all but one site underwent at least one documented anemometer height (AH) relocation (as well as instrument change). The majority (25 of 31 sites) have undergone at least three such occurrences. Since wind speed scales non-linearly with height, the AH changes will introduce discontinuities to a wind speed time series. Within the Canadian surface wind observations, the AH changes represent a number of potentially significant inhomogeneities in the data. This issue was addressed in Wan et al. (2010); wherein historical surface wind speed hourly observations were adjusted to the standard 10 meter anemometer height using station metadata and a logarithmic profile.

Wan et al. (2010) found that AH changes were generally the largest, but not sole, source of discontinuities in the observational records of surface winds. Instrument changes and horizontal relocations also represented significant non-stationarities, larger at some locations than those from AH changes. Non-stationarities may also exist in the data records from causes that do not stem directly from the data acquisition faculties. Changes in a station's surrounding vegetation, for instance, will alter the local roughness length (albeit likely small at airports). A change in the surrounding structural environment may also alter the path and strength of the surface flow. This would be particularly true of the only observational record in this study that did not undergo any station relocations or instrument changes; the Edmonton City Center Airport. To minimize these effects, Wan et al. (2010) have also provided a dataset of monthly mean wind speeds computed from the adjusted hourly observations and homogenized using statistical methods (Wang, 2008) and a grid of homogenized sea level pressure readings. The manner in which these data are homogenized is important; it would be inappropriate to assess downscaling predictability using data that underwent a statistical correction involving large-scale free-tropospheric reanalysis data. Since reanalysis data was not used in the homogenization methodology, the homogenized data could therefore be used to assess predictability with the reanalyses products without concerns of spurious prediction skill.

Unfortunately, the homogenized dataset is available only as monthly speeds, lacking wind direction as well as hourly observations renders it unsuitable for this study. Instead, the hourly data used to form the predictands were taken from two sources. Unadjusted hourly data from the Meteorological Service of Canada in situ station records were used to provide the wind directions. These observations were retrieved directly from the Environment Canada Weather Office Climate Data Online service at *climate.weatheroffice.gc.ca/climateData/canada_e.html* (Canada, 2011). The hourly observations contain the average of the direction and speed of the wind during the two-minute period ending at the hour of observation. Further information regarding the wind measurement instrumentation and observation methods is available in the "Manual of surface weather observations"(Canada, 1977). To minimize the effect of station relocations, the magnitude of the surface wind was taken from the adjusted historical hourly surface wind speeds (Wan et al., 2010). The slight disadvantage of using the adjusted dataset is a restriction of the pool of available sites (with ten in Alberta, eight in Saskatchewan, and nineteen in Ontario) and lack of data beyond

2006. These were retrieved directly from from the Environment Canada Adjusted and Homogenized Canadian Climate Data (AHCCD) service at *http://ec.gc.ca/dccha-ahccd/default.asp?lang=en&n=552AFB3E-1.*

2.1.2 Predictors

The statistical model's predictors were constructed from large-scale fields of retrospective analysis (reanalysis) products. Reanalysis products are four-dimensional datasets representing historical states of the earth's atmosphere. The data is created through the assimilation of observations with a numerical weather prediction model. Values of meteorological variables from initial outputs of the the numerical model are compared with observations, the model bias is calculated (where observations are available), and this is used to nudge the numerical model during a subsequent iteration (Uppala et al., 2005). In this way, the simulated representation of the earth's atmosphere follows the observational records. The reanalysis dataset describes surface level meteorological variables such as precipitation, temperature, and surface winds, as well as meteorological variables at various pressure levels throughout the troposphere and lower stratosphere.

In this study, two reanalysis products were used to produce two distinct sets of predictors: the North American Centers for Environmental Predictions (NCEP) North American Regional Reanalysis (NARR) product (Mesinger et al., 2006), and the NCEP / National Center for Atmospheric Research (NCAR) Reanalysis I (R1) product (Kalnay et al., 1996; White et al., 2001). Both reanalysis products were provided by the NOAA/OAR/ESRL Physical Sciences Division, Boulder, Colorado, USA from their FTP site at *ftp://ftp.cdc.noaa.gov/Datasets/*. The NARR product is a regional dataset covering the North American region on a 32 km by 32 km equal area grid with 4 hour timesteps, spanning 1979 to the present. In contrast the R1 product is a global dataset, with a 2.5^o by 2.5^o grid size and a 6 hour timestep, spanning 1948 to the present. The study considered climate data from the regional reanalysis product to assess the extent to which predictive information occurs on scales not resolved in the global reanalysis dataset. Conversely, the longer duration global reanalysis product provides an opportunity to assess predictability over a longer time period.

With the improvements of the data assimilation of NARR over the R1 product, as well as the greater resolution and updated dynamics, it is no surprise that NARR more closely matches observations. Studies have shown that NARR's near surface (i.e. 2 m and 10 m) values of precipitation (e.g Bukovsky and Karoly, 2007), as well as wind and temperature (e.g. section 3, Kanamaru and Kanamitsu, 2007) show greater accuracy than R1. However, the strongest gains in the accuracy (both RMS and bias) of the vector winds were found in the mid and upper tropospheric winds (Mesinger et al., 2006). It is of interest whether the improved accuracy and finer model resolution translates to additional predictive information. The extent of the predictive information in the two reanalysis products could certainly be quantified by performing and contrasting predictions from each product. A simpler comparison follows from examining maps of the correlations of the flow aloft from each product with the surface winds at any location.

The strength and horizontal extent of the correlations of the monthly-mean DJF wind speed observed at the Lethbridge airport and the zonal flow at the 500 hPa level in the NARR reanalysis (left), and in the NCEP reanalysis I (right) are displayed in Figure 2.2. There are two clear, but superficial, differences evident in these plots: the finer resolution of NARR, and the disparate grid projections. The predictive information is common between the reanalysis products, as evidenced by the spatial pattern and strength of the correlations. With both products, positive anomalies in the mean wind speed at Lethbridge are associated with anomalies in the zonal winds that are positive over the north-eastern Pacific Ocean and north-western continental North America, and negative over the southwestern continental North America and the neighbouring region of the Pacific Ocean. A similar (yet weaker) relationship with the zonal winds over the eastern coast of North America and the Atlantic Ocean is also seen. Good agreement between reanalysis products was also observed at other Alberta, Saskatchewan, and Ontario sites, as well as during other calendar seasons.

Evidently, despite the improved accuracy of the surface and mid-tropospheric variables in the NARR reanalysis, the spatial and temporal extent of the predictive information in the flow aloft is such that there is no added gain from the finer resolution. The strength of the correlation suggests that the tropospheric flows prescribed by the NCEP R1 compare quite favourably with the flows derived by NARR. In this case, there is no penalty in foregoing finer resolution for longer duration when select-

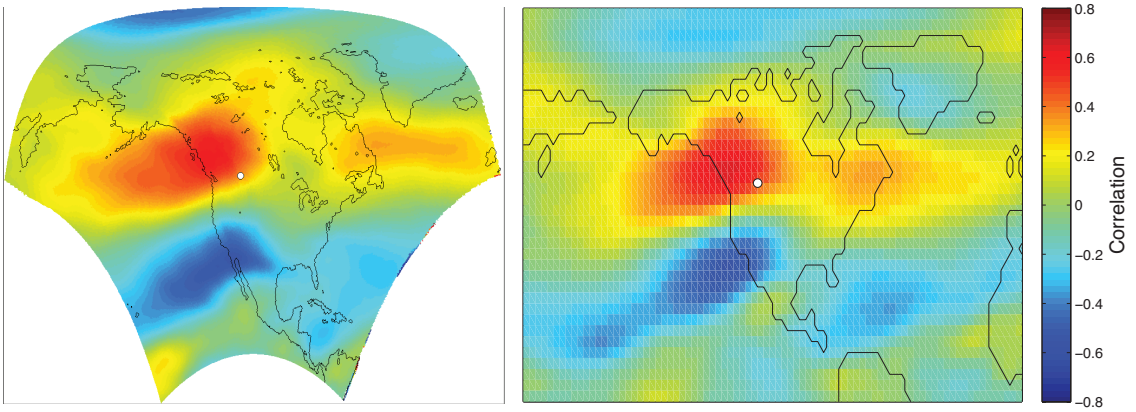

Figure 2.2: The correlations of wintertime monthly mean wind speeds at Lethbridge AB and the mean zonal flow at 500 hPa: as computed with data from the NARR reanalysis (left), and the NCEP reanalysis I (right). The difference in the shapes of the domains is due to the different mapping projections used by the reanalysis products. NARR's output is on a Lambert Conformal Conic projection; the area between grid points is generally preserved. NCEP-R1 maps to a Cylindrical Equidistant projection; here the latitudinal and longitudinal spacing between grid points is constant.

ing the predictors, and for the remainder of this study predictors will be derived from the R1 product.

2.2 Methods

2.2.1 The Statistical Model

This study employs a multivariate linear regression model to capture the relationship between the information carried in the flow aloft and the local surface winds. In very general terms, linear regression models a linear relationship between two variables. With multivariate regression, the predictand (variable), $\mathbf{y}$, is modelled as a sum of individual linear relationships with each of the predictors (variables), $\mathbf{x}$:

$$\hat{\mathbf{y}} = \mathbf{xb}, \tag{2.1}$$

or, for k predictors

$$\begin{bmatrix} \hat{y_1} \\ \hat{y_2} \\ \vdots \\ \hat{y_n} \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,k} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} \tag{2.2}$$

The predictand is defined as the sum of the model prediction, $\hat{\mathbf{y}}$, and the prediction error, ϵ (often referred to as the residuals),

$$\mathbf{y} = \mathbf{xb} + \epsilon. \tag{2.3}$$

The regression technique seeks linear coefficients, b_i, that minimize the sum of the square of the error between the predictions, $\hat{\mathbf{y}}$, and observations of $\mathbf{y}$. This result is obtained when

$$\mathbf{b} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y}. \tag{2.4}$$

The variance of the predictand ($std^2(\mathbf{y})$) is a sum of the variance of the model predictions ($std^2(\hat{\mathbf{y}})$) and the variance of the residuals ($std^2(\epsilon)$). By definition, variance is positive definite. The performance of the regression can be characterised by how much of the variance of $\mathbf{y}$ is captured by that of $\hat{\mathbf{y}}$. The common metric for this is r^2, the coeffecient of determination:

$$r^2 = \left(\frac{std(\hat{\mathbf{y}})}{std(\mathbf{y})}\right)^2 \tag{2.5}$$

$$= 1 - \left(\frac{std(\epsilon)}{std(\mathbf{y})}\right)^2. \tag{2.6}$$

The coefficient of determination (r^2) is bounded by 0 and 1, where the upper bound signifies a perfect prediction and the lower bound is a meaningless prediction. Herein, r^2 will be the metric used to assess prediction skill.(Wilks, 1995)

Any statistical model based on minimizing an error will try to get the best possible parameter estimates from the data used for the estimation procedure. As such, random synchronizations in fluctuations between predictor(s) and a predictand will be

treated as signal, and the estimated model parameters will reflect this. When applied to a new set of data, these parameters will be sub-optimal. This problem is known as overfitting of a statistical model, and generally becomes worse as the number of model parameters increases. For this reason, care must be made when choosing the number of model parameters as well as the means in which the parameters are calculated.

Since linear regression requires a relatively low number of model parameters, the risks of overfitting the model are lower than with a more complex but possibly better performing model. This fact is particularly significant as statistical downscaling is dependent on historical observations, where datasets may be small, to characterize the relationship with the predictor(s). A low number of statistical degrees of freedom increases the risk of overfitting with more complex models. Furthermore, linear regression is computationally efficient, widely used, and well understood. As such, it is a useful tool for a first assessment of the predictability of surface winds.

The disadvantages of linear regression are also a result of its simplicity. At the heart of this technique is the assumption that the relationship between predictor(s) and predictand is linear. As well, this manner of regression does not make adjustments for instances of non-Gaussianity of the errors. Consequently, it will typically describe only a portion of y's variance if the relationship between it and observations of x are non-linear (or more generally non-Gaussian). Given the simplicity of the model used, the prediction skills found in this study may be interpreted as a floor for potential prediction skill, where higher values may be achieved with more complex statistical downscaling techniques.

Cross-Validation

Cross-validation schemes are a necessary step to obtain a robust statistical downscaling models. The common tenet among these schemes is the statistical independence of the information used to train the model (i.e. calculate the regression model parameters) and the information that is predicted. For daily, monthly-, and seasonally-averaged surface wind quantities, year-to-year correlations are expected to be small. Consequently, in this study each year was predicted individually. The year being predicted was "bagged" (i.e. withheld) and data from all other years was used to calculate the regression model parameters. The specified year was then predicted from

the calculated coefficients and the predictor values. The regression model parameters are then discarded and the process repeated for each successive year. This method is preferred over methods which randomly sample the values to "bag", as such a method omits a portion with a duration greater than the autocorrelation length of the surface winds. In this way the study sought to ensure a robust statistical downscaling model.

2.2.2 Constructing the Predictands

In contrast to past studies of statistical downscaling surface wind speeds and vector components (e.g. Salameh et al., 2009; Goubanova et al., 2011; Najac et al., 2009), this study did not focus solely on either the wind speed or the zonal and meridional flow. Following the approach of van der Kamp et al. (2011), the wind speed as well as a full 360^o array of vector projections were predicted. Projections of the wind vector were made at ten degree increments spanning 170^o, yielding 18 wind vector projections per site (by construction, projections are equivalent in directions 180^o apart). To avoid complications arising from potential seasonal non-stationarities in the relationship between surface flow and the statistics of the troposphere, each calendar season (DJF, MAM, JJA, SON) was independently predicted.

Two specific statistical features of the surface winds were calculated: the mean and the standard deviation. Wind speed readings of zero were omitted from the calculations of these statistical features. This was done for two reasons: first, to maintain consistency with previous studies (e.g. van der Kamp et al., 2011); and second, measurements of zero are not expected to be in reality speeds of zero, but represent flow below the lower limit of detection of 2 km hr^{-1}. Sensitivity analyses (not shown) demonstrated that results of these analyses are not sensitive to the inclusion or exclusion of these zero wind speed readings.

The means and standard deviations were calculated at daily, monthly, and seasonal averaging timescales. There are three motivations for predicting both means and standard deviations of surface winds.

i. At any particular averaging timescale, predicting both quantities allows the information aloft to be assessed for its ability to predict surface wind variability occurring on sub-averaging and on super-averaging timescales.

ii. For monthly and seasonal averages, the sub-averaging timescale variability captures the majority of the variable's variance (which is concentrated on synoptic timescales) and is of interest in a number of applications, e.g. wind power calculations, meteorological support for forward energy trading, surface flux calculations, architectural load calculations, among others.

iii. Because the transformation from vector wind components to wind speed is nonlinear, the mean wind speed is a function of both the means and standard deviations of the vector wind components.

This study assessed the predictability of both vector and scalar winds. The study's vector wind predictands were,

$$\begin{aligned} \overline{\tilde{u} \cdot \hat{e}} &= \text{mean of a wind vector projection along the unit vector } \hat{e}, \\ \sigma_{\tilde{u} \cdot \hat{e}} &= \text{standard deviation of a wind vector projection.} \end{aligned}$$

The study's scalar wind predictands were,

$$\begin{aligned} \overline{w} &= \text{mean wind speed}, \\ \sigma_w &= \text{standard deviation of the wind speed.} \end{aligned}$$

Two additional vector wind predictands, associated with the idealised model used to relate the predictability of vector and scalar wind statistics (Chapter 4), were calculated. They were,

$$\begin{aligned} \mu &- \text{the magnitude of orthogonal mean vector winds, } \mu = \sqrt{\bar{u}^2 + \bar{v}^2}, \\ \sigma &- \text{the isotropic fluctuations of the vector winds, } \sigma = \sqrt{\frac{1}{2}\left(\sigma_u^2 + \sigma_v^2\right)}. \end{aligned}$$

2.2.3 Selecting the Predictor Variables

The reanalysis products offer a suite of potential predictors at the surface level and at pressure levels throughout the troposphere. A short list of predictor variables that have consistently produced meaningful prediction skill across previous studies (e.g. Salameh et al., 2009; Najac et al., 2009; van der Kamp et al., 2011; Curry et al., 2011; Davy et al., 2010) is:

1. wind speed - W
2. zonal flow - U
3. meridional flow - V
4. geopotential height - Z_g
5. temperature - T

There is also an observational and theoretical basis to expect these variables to have a strong relationship with the surface winds, and hence use them to drive the statistical model. On synoptic and longer timescales, the primary dynamic processes contributing to surface variability have structure throughout the troposphere. Furthermore, large-scale balances couple atmospheric mass distribution, thermodynamic structure, and flow.

To assess which of the listed predictors to include in the model, the correlation of the local observations and the predictors at pressure levels spanning the troposphere was mapped. These correlation maps are themselves a form of prediction and hence a measure of attainable prediction skill. A high correlation with the predictor variable indicates a strong linear relationship between the local surface flow and the information contained in the flow aloft. However, correlation fields are not a cross-validated measure of prediction skill, so the correlation corresponds to a potential prediction skill ceiling. The spatial scale of the predictive information is also a relevant measure. For a predictor to be useful in statistical downscaling, its correlation structure with the predictand must be on horizontal scales that can be well-resolved by General Circulaton Models. As a result, those predictors that exhibited large areas with predictive information throughout the lower to mid troposphere were retained, those that did not were discarded.

An illustration of the horizontal extent and location of predictive information in the predictor variables relevant to mean wind variability at Lethbridge is shown in Figure 2.3. Here the correlations were calculated over DJF seasonal means from 1953 to 2006. The figure illustrates the correlation fields of the predictor variables at the 600 hPa level with the surface wind's mean speed and mean zonal and meridional components. The spatial extent of the predictive information contained in the seasonal statistics of the considered predictors is clearly demonstrated.

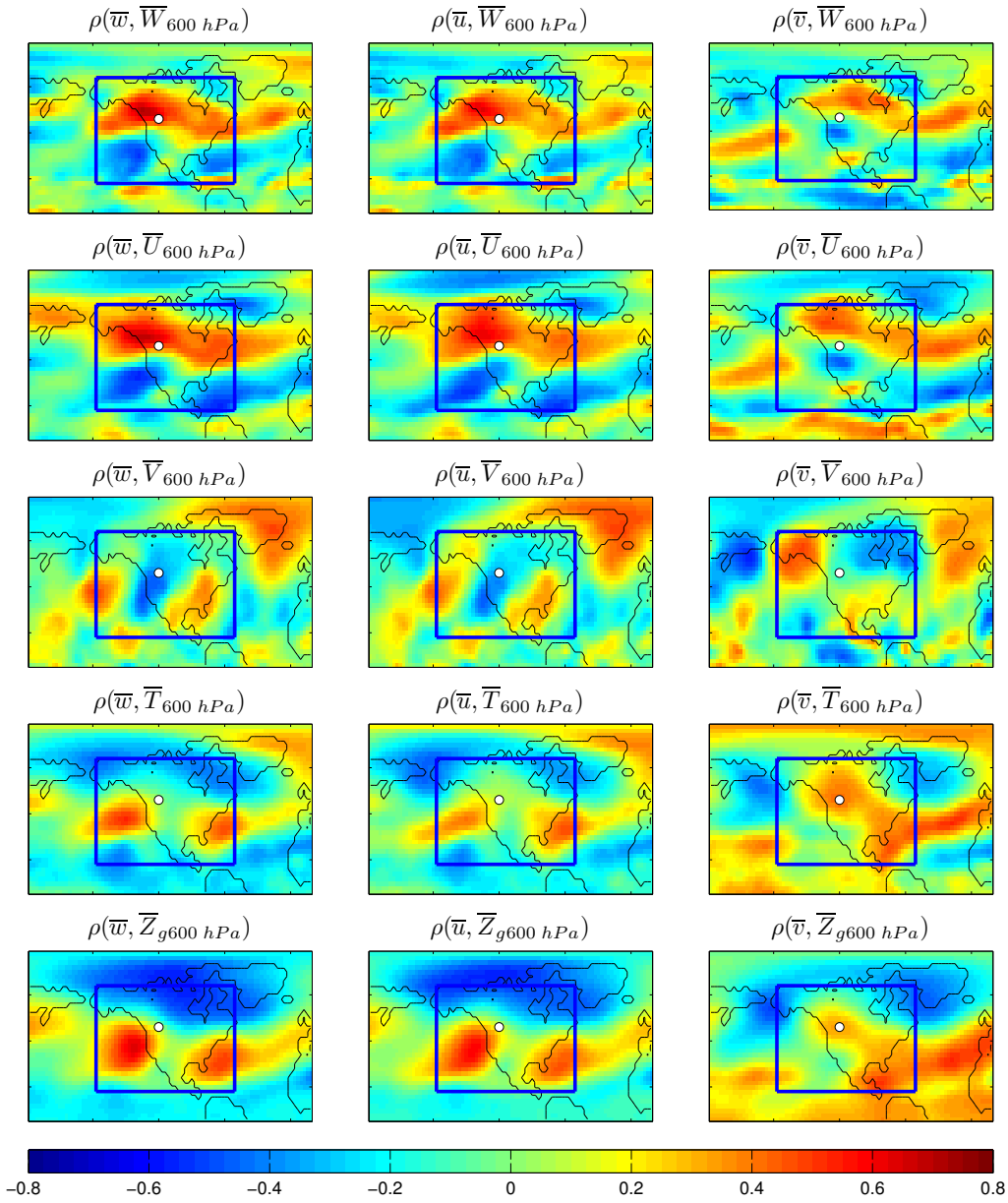

Figure 2.3: The correlation of the observed DJF seasonal means at Lethbridge (white circle) and the 600 hPa predictor variables. The predictands are grouped by column: mean wind speed - $\overline{w}$ (left column), mean zonal flow - $\bar{u}$ (middle column), mean meridional flow- $\bar{v}$ (right column); and the predictor variables are organized by row: mean wind speed aloft - $\overline{W}$ (top row), mean zonal flow aloft $\overline{U}$(second row), mean meridional flow aloft - $\overline{V}$ (third row), mean temperature aloft $\overline{T}$ (fourth row), and mean geopotential height - $\overline{Z}_g$ (bottom row).

The horizontal scale and location of predictive information at monthly and daily averaging timescales is shown in Figures 2.4 and 2.5. The spatial extent and location of the predictive information aloft appears to have a strong sensitivity to the averaging timescale. The predictive information aloft is found on increasingly larger spatial scales as the averaging timescale increases. The location of the maximum correlation values are generally far removed from the predictand position, increasingly so as the averaging timescale increases.

These correlation plots have a consistent physical interpretation. Consider the correlation of the seasonal wintertime mean wind speed at Lethbridge and the mean zonal flow at 600hPa (Figure 2.3; left column, second row). We find that wind speeds at Lethbridge are larger than average when the mid-latitude jet stream shifts poleward. Conversely, when the zonal jet shifts equatorward, the winds at Lethbridge decrease. The correlation plots of T and Z_g and mean wind speed at Lethbridge (Figure **??**; left column, fourth and fifth row) demonstrate that stronger than average surface mean wind speeds occur when the mid- to upper-latitude meridional temperature gradient and hence meridional geopotential height gradient is strengthened.

The dependence of the predictive information aloft on the averaging timescale clearly discourages the use of SD methods utilizing a nearest neighbour (or grid point) predictors approach (i.e. Pryor et al., 2005; Davy et al., 2010). The separation between the location of maximum correlation and the predictand in Figures 2.3, 2.4, and 2.5 illustrates that nearest-neighbours will not be the best available, nor necessarily good predictors. This is particularly the case with longer averaging timescales.

As a consequence of the strength and spatial scale of their correlation values, the daily, monthly, and seasonal means of the predictor variables listed above were used to form the study's predictors. The information contained in the sub-averaging timescale standard deviations of these free tropospheric variables was also evaluated. Due to a lack of any evidence of large-scale predictive information in these standard deviation fields, they were omitted from further consideration as predictors.

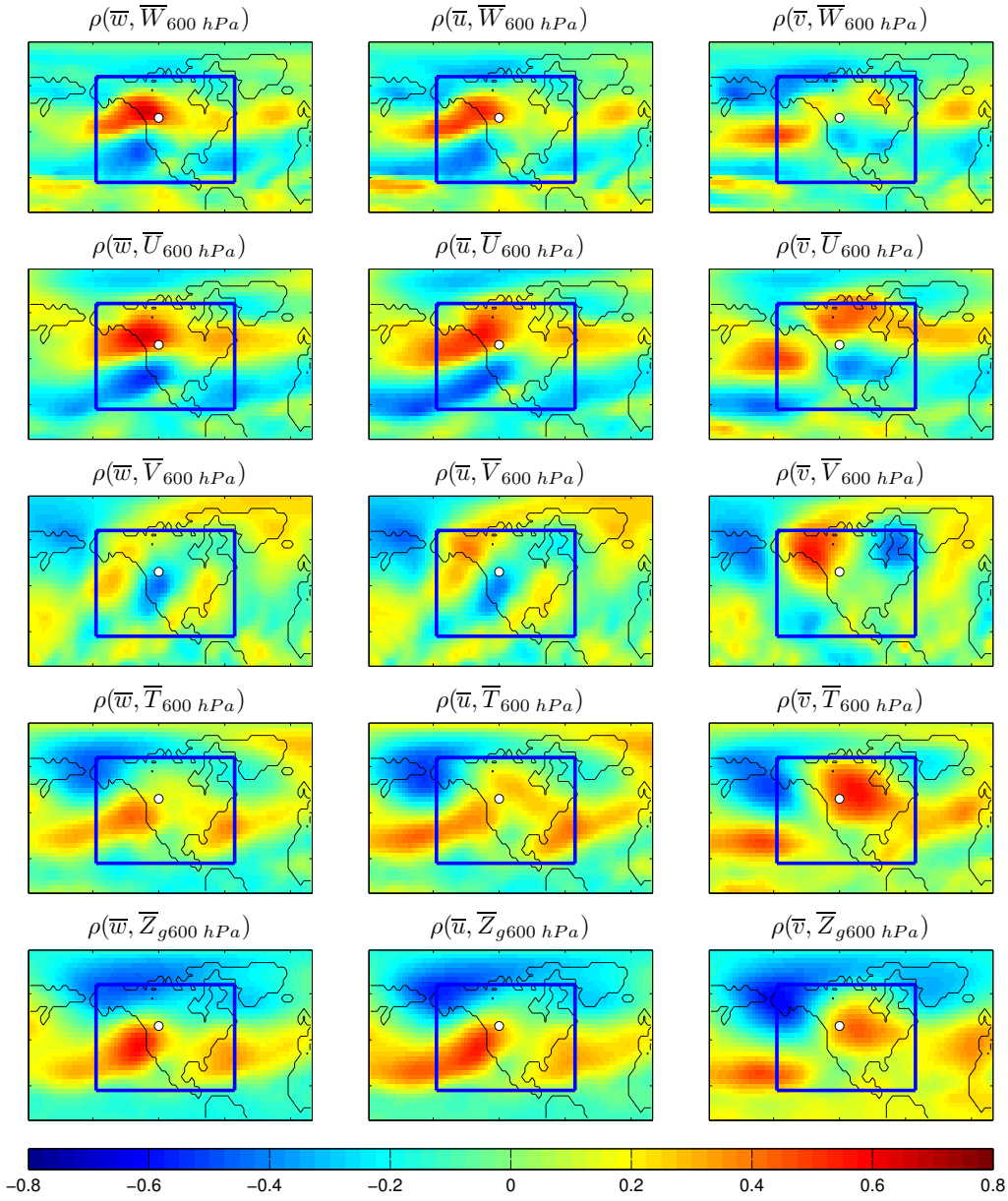

Figure 2.4: As with Figure 2.3 but with monthly means.

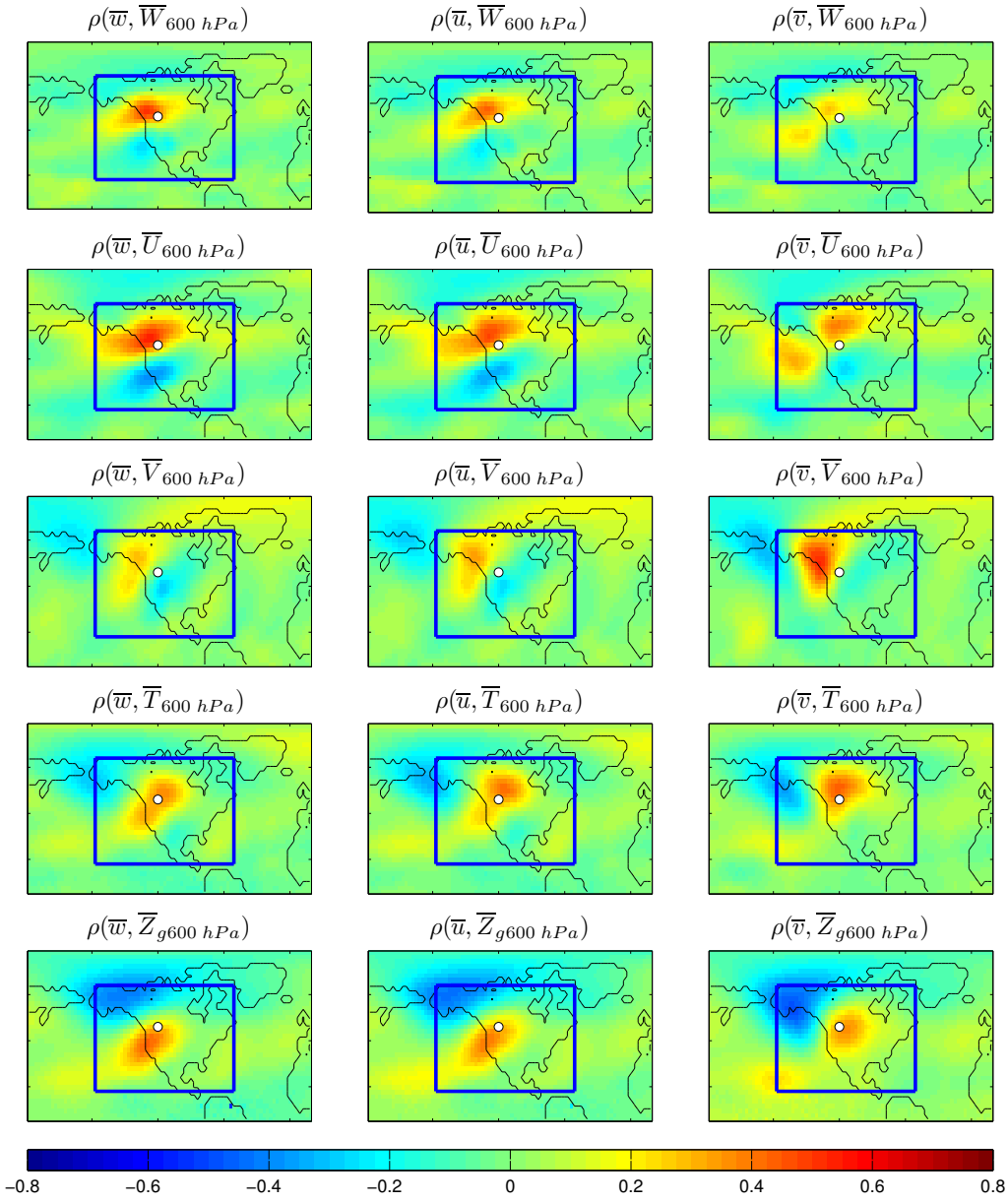

Figure 2.5: As with Figure 2.3 but with daily means.

Creating the Predictor Time Series

The statistical model used in this study formed predictions from multiple time series obtained from a set of predictor variables. As the number of predictors increases, the chances of being overfit increase. For successful predictions, it is necessary to select a minimum number of time series that contain a maximum amount of predictive information.

With any predictor variable, the larger the horizontal scale of predictive information, the larger the degree of shared information (i.e. redundancy) across the time series of the predictor variable at different grid points. It is apparent from the correlation maps (Figures 2.3, 2.4, and 2.5) that the predictive information in the flow aloft is found on large scales, particularly on seasonal and monthly averaging timescales. There is also redundancy expected across predictors coupled by large-scale force balances. The bottom two rows of Figures 2.3, 2.4, and 2.5 illustrate the predictive information contained in the mean air temperature and mean geopotential height predictors; note their similarity. From these plots, we see the number of linearly independent variables carried in the predictor fields is substantially smaller than the raw dimensions of the predictor set (given by the number of grid points multiplied by the number of fields considered).

To reduce the dimensionality of the predictors, an Empirical Orthogonal Function (EOF) decomposition is used. The central purpose of EOF analysis (also known as principal component analysis) is the reduction of the dimensionality of a dataset of interrelated variables while retaining as much of the dataset's variance as possible (Wilks, 1995). For variables with well behaved correlation structures, such as those in Figures 2.3, 2.4, and 2.5, the leading EOFs contain the large-scale structures, and higher order EOFs have increasingly smaller scales (Buell, 1979). By retaining only a subset of the leading-order EOFs, the variance of the predictor fields on larger scales is retained. Accordingly, EOF analysis was performed to produce a set of linearly independent variables containing the large-scale predictive information in the flow aloft.

For each of the four calendar seasons, the five predictor variables were decomposed into their EOFs, creating sets of orthogonal spatial maps of the modes of variance with corresponding sets of time series. The spatial domain over which the EOFs

were calculated are shown in Figure 2.3; the same domain was used at all averaging timescales. In principle, the EOFs of individual fields could be used as statistical downscaling predictors. However, to account for the clear covariance among the predictors, combined EOFs were calculated. Each EOF field was normalized by the square-root of its spatial mean variance and these were concatenated to make a single predictor vector time series. The normalization accounts for the fact that the predictor variables have differing variability and units, and ensures that each predictor variable contributes the same variance to the combined field. A second iteration of EOF analysis was then carried out on the combined field. This created a single set of predictors that most economically described the dominant statistical modes of variance among the predictor variables.

Tables 2.1, 2.2, and 2.3 present the percentage of total variance explained by the individual EOF and combined-EOF wintertime predictor fields on seasonal-, monthly-, and daily-averaging timescales respectively. The rationale in choosing the number of predictors listed in these tables will be discussed in Chapter 3.

Table 2.1: The percentage of variance explained (σ^2) by the leading individual EOF and combined-EOF predictor fields on a seasonal averaging timescale, as well as the percentage of variance explained by the number of predictors (#P) used in the SD model.

Var.	EOF_1	EOF_2	EOF_3	EOF_4	EOF_5	EOF_6	$\sum_{i=1}^{6} \sigma_i^2$	#P	$\sum_{i=1}^{\#P} \sigma_i^2$
W	31.6	15.5	9.9	7.3	5.4	4.3	74.0	6	74.0
U	36.1	16.1	12.4	7.5	6.1	3.8	82.0	6	82.0
V	28.7	17.1	12.6	8.6	6.3	5.0	78.3	6	78.3
T	37.8	18.3	14.8	5.1	4.6	3.3	83.9	6	83.9
Z_g	40.7	18.6	14.1	7.3	5.7	3.0	89.5	6	89.5
Combined	29.2	16.3	12.4	8.1	5.6	3.6	75.2	6	75.2

There are two key points to note from these tables. First, the spatial modes are ranked by the amount of variance they carry in the data, with the leading modes being large-scale structures and carrying the largest variance. It is these leading EOFs that formed the predictors. Second, that the combined EOF is always more efficient than the separate EOFs at describing the total variance of the five predictor variables. Note

Table 2.2: As with Table 2.1 for a monthly averaging timescale.

Var.	EOF_1	EOF_2	EOF_3	EOF_4	EOF_5	EOF_6	$\sum_{i=1}^{6} \sigma_i^2$	#P	$\sum_{i=1}^{\#P} \sigma_i^2$
W	23.9	13.0	10.4	8.1	6.6	5.0	67.0	10	77.3
U	28.1	15.2	13.3	8.8	7.4	4.7	77.6	10	87.2
V	26.0	19.6	11.1	9.1	6.0	4.3	76.1	10	87.3
T	33.0	17.3	12.8	8.7	5.7	3.9	81.3	10	91.3
Z_g	31.0	20.1	14.9	8.6	7.6	4.2	86.3	10	94.7
Combined	22.6	15.3	11.3	8.8	6.1	5.3	69.5	10	81.0

Table 2.3: As with Table 2.1 for a daily averaging timescale.

Var.	EOF_1	EOF_2	EOF_3	EOF_4	EOF_5	EOF_6	$\sum_{i=1}^{6} \sigma_i^2$	#P	$\sum_{i=1}^{\#P} \sigma_i^2$
W	9.2	8.0	7.2	6.2	5.0	3.7	39.2	125	99.3
U	13.2	11.3	9.4	8.3	6.1	5.2	53.4	125	99.8
V	14.9	13.3	9.7	7.7	5.6	5.1	56.4	125	99.8
T	19.2	15.0	10.4	8.0	6.1	4.4	63.1	125	99.9
Z_g	21.8	19.1	12.4	8.2	7.7	5.1	74.3	125	100.0
Combined	12.0	10.1	8.0	6.6	5.7	4.5	46.8	125	95.6

that because the EOFs of the individual fields describe only a portion of that field's variance, the portion of the total variance of the predictor variables would be one fifth of the values quoted in the table. For instance, considering monthly timescale predictors, the leading two combined EOFs explain $\approx 38\%$ of the total variance of W, U, V, T, and Z_g. On the other hand, 7 individual EOFs (2 eigenvectors of V and Z_g and 1 of W, U, and T) are required to explain a nearly equal proportion of the total variance. By optimally packaging the data, the combined-EOF components allow the statistical model to map a greater amount of the predictive information aloft to the predictands with a considerably reduced risk of overfitting.

Since leading EOFs are large-scale structures, the representation of localized features requires many more EOFs than the representation of unlocalized features. Consequently, localized (or "noisy") predictor variables such as vertical advection were omitted from the predictor construction as their predictive information is not well

resolved by a small set of EOF modes. Furthermore, it is expected that more EOF modes will be needed to capture the correlation structure on daily timescales than on monthly or seasonal timescales. Fortunately, the number of statistical degrees of freedom is correspondingly higher on shorter averaging timescales, allowing larger numbers of predictors to be used.

Chapter 3

Assessing Predictability

3.1 Quantifying the Predictive Information Aloft

Having identified the predictor variables that capture information relevant to surface winds in the tropospheric circulation, it was necessary to determine the vertical extent of the predictive information in the flow. Studies on statistical downscaling of surface winds have used predictive information at pressure levels ranging from the surface to the mid-troposphere (e.g. Salameh et al., 2009; Najac et al., 2009; van der Kamp et al., 2011; Curry et al., 2011; Davy et al., 2010). The vertical extent of the predictive information in the reanalysis products was evaluated explicitly by making predictions at all available pressure levels.

Figure 3.1 illustrates the r^2 prediction skill of DJF predictions at eight Alberta sites as a function of the predictors' pressure level. The solid curves illustrate the predictability of the monthly means of the wind speed, the zonal wind, and the meridional wind, and dashed curves illustrate the predictability of the sub-monthly standard deviations of these quantities. The range of predictability across the predictands is fairly small ($\Delta r^2 \simeq 0.2$) at Grande Prairie and Peace River; on the other hand, it is much larger at Edmonton and Lethbridge. In general, we see from this Figure that at these sites

i. the means are better predicted than the standard devations,

ii. the means of the meridional wind are generally the best predicted quantity,

iii. the predictability of the mean wind speed is worse than that of the mean vector wind components and in most cases no better than the predictability of the standard deviation of the vector components.

We will discuss these general results in more detail in later sections.

We also see from this Figure that in general prediction skills are not strongly dependent on the pressure level of the predictors. In particular, whether predictions are good or quite poor, predictability is largely insensitive to pressure level between 800 hPa and 400 hPa. Because it is expected that free tropospheric winds will be better represented by large-scale climate models than winds in the boundary layer, predictor variables from the 600 hPa reanalysis fields were used to drive the predictions.

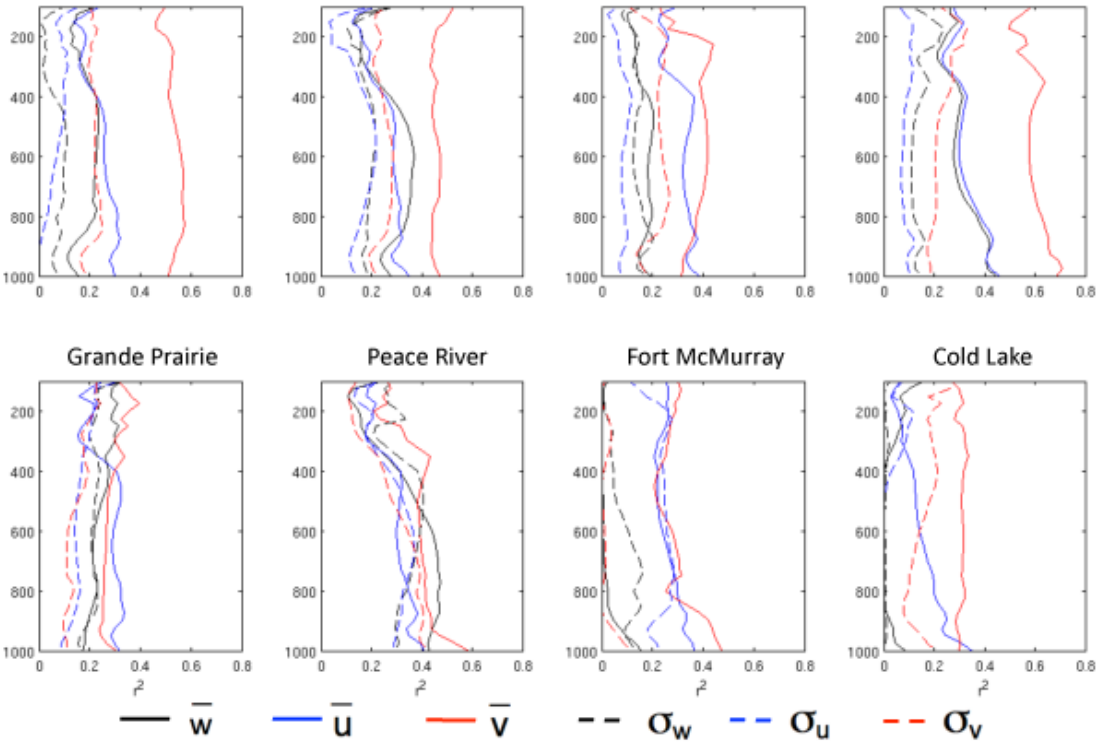

Figure 3.1: The dependence of the cross-validated r^2 monthly prediction skill on the atmospheric pressure level of the statistical downscaling predictors.

3.1.1 The Number of Predictors Used

The use of cross-validation to estimate model parameters will reduce, but not eliminate, overfitting. With any statistical prediction model there must also be a balance between minimizing the mapping of spurious information to the model parameters and the capability to model complex statistical relationships. For the observations under consideration, the nominal numbers of statistical degrees of freedom at seasonal, monthly, and daily averaging timescales are 54, 162, and approximately 4800 respectively. However, as a result of autocorrelation in both predictors and predictands, the actual statistical degrees of freedom will be lower. Over-fitting becomes a greater concern at longer averaging timescales, as the number of statistical degrees of freedom available decreases. Similar to Monahan (2012a), a qualititative method was used to select an appropriate number of predictors.

Eight sites were chosen at random, and monthly-timescale predictions were made with increasing numbers of predictors (Figure 3.2). Beyond the very first few predictors, the cross-validated prediction skill curves change gradually with predictor number, so predictability is insensitive to small changes in the number of predictors. The r^2 curves saturate because they are constructed using cross-validated predictions; for non-cross-validated predictions, prediction r^2 will increase monotonically with predictor number. When the number of predictors approaches the number of statistical degrees of freedom, our model becomes grossly overfit. At this point, r^2 approaches zero as the model fails completely when it is applied to new data. The grey curve in Figure 3.2 represents the cumulative variance in the combined predictor field explained by the predictors. Comparing the grey curve and the saturation point of the r^2 curves, we see that beyond the predictors containing the large-scale structure aloft (the leading 10 or so EOFs), the additional $\simeq 30\%$ of the variance present in the flow aloft does not contain robust predictive information for the monthly statistics of surface winds. Similar calculations (not shown) were used to determine the number of predictors to be used on daily and seasonal timescales.

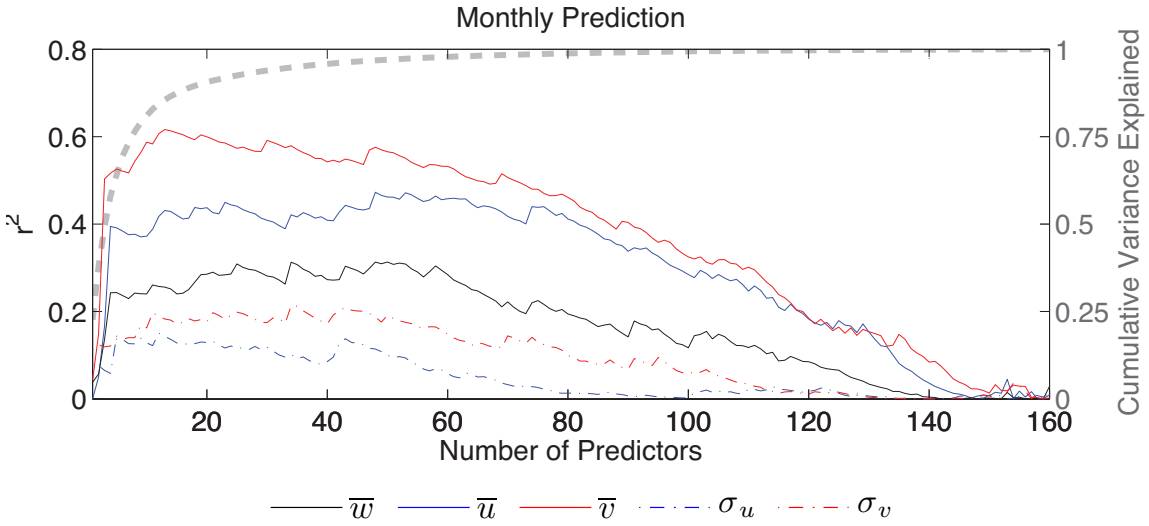

Figure 3.2: The cross-validated r^2 prediction skill (solid and dot-dashed lines) and the cumulative variance of the reanalysis data explained (dashed line) as a function of the number of model predictors.

3.2 Prediction Results

Having created predictors for each of the three-month calendar seasons (DJF, MAM, JJA, and SON), and identified on which vertical levels the predictive information was carried, predictions of historical wind observations were made. This was done using predictors averaged on daily, monthly, and seasonal timescales. A typical representation of the cross-validated r^2 prediction skill of vector and scalar predictands is shown in Figure 3.3 using the DJF Lethbridge results. These polar plots show four different prediction skills, one for each of the two vector component predictands (mean and standard deviation; red curves) and the two wind speed predictands (mean and standard deviation; blue curves). With both the vector components and the wind speed, the dashed lines show the prediction skill achieved with the standard deviations, and the solid line shows that achieved with the means. The outer black circle denotes a reference r^2 of 0.6. Such polar plots were computed for each station, season, and averaging timescale.

These plots clearly demonstrate an anisotropy of prediction skill for vector quantities. At Lethbridge (Figure 3.3), there are clearly a maximum and a minimum in the predictability of the vector wind predictands. The statistics of the south-westerly

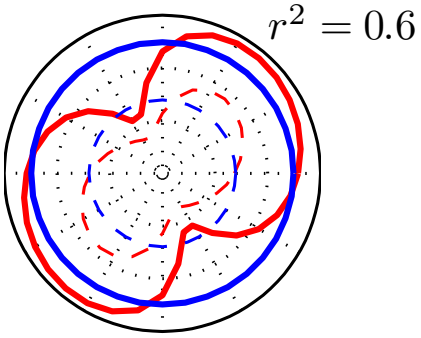

Figure 3.3: A representative plot of the cross-validated r^2 prediction skill for the vector and scalar predictands. The blue circles denote the prediction skill achieved with the two scalar predictands ($\overline{w}$ and σ_w). The red polar plots illustrate the prediction skill achieved with the two vector predictands ($\overline{\tilde{u} \cdot \hat{e}}$ and $\sigma_{\tilde{u} \cdot \hat{e}}$). The dashed and solid linetypes denote the predictability of the standard deviations and means respectively. The outer black circle denotes a cross-validated r^2 prediction skill of 0.6. Each successive inner dotted circle corresponds to a r^2 increment of 0.2.

(and north-easterly) vector winds are best predicted; while predictability of the vector winds is markedly reduced for north-westerly and south-easterly projections. Such anisotropy was first noted by van der Kamp et al. (2011), and is ubiquitous to the prediction results. These plots also illustrate the relative predictability of the scalar and vector wind predictands. At Lethbridge, predictability of the monthly-mean scalar wind speed is comparable to that of the best-predicted monthly-mean vector wind component. In all directions, the monthly mean vector wind component is better predicted than its sub-monthly standard deviation.

Polar r^2 prediction skill plots for the monthly mean DJF winds at the Alberta and Saskatchewan sites are shown in Figure 3.4. These plots follow the same format as Figure 3.3, except black arcs have replaced the solid black circle representing a reference cross-validated r^2 prediction skill of 0.6. The plot demonstrates considerable variability between stations in the magnitude and anisotropy of predictability. There are sites that appear to have similar predictability features, such as the two pairs of northern Alberta sites; even among these sites, striking differences are readily apparent upon close scrutiny.

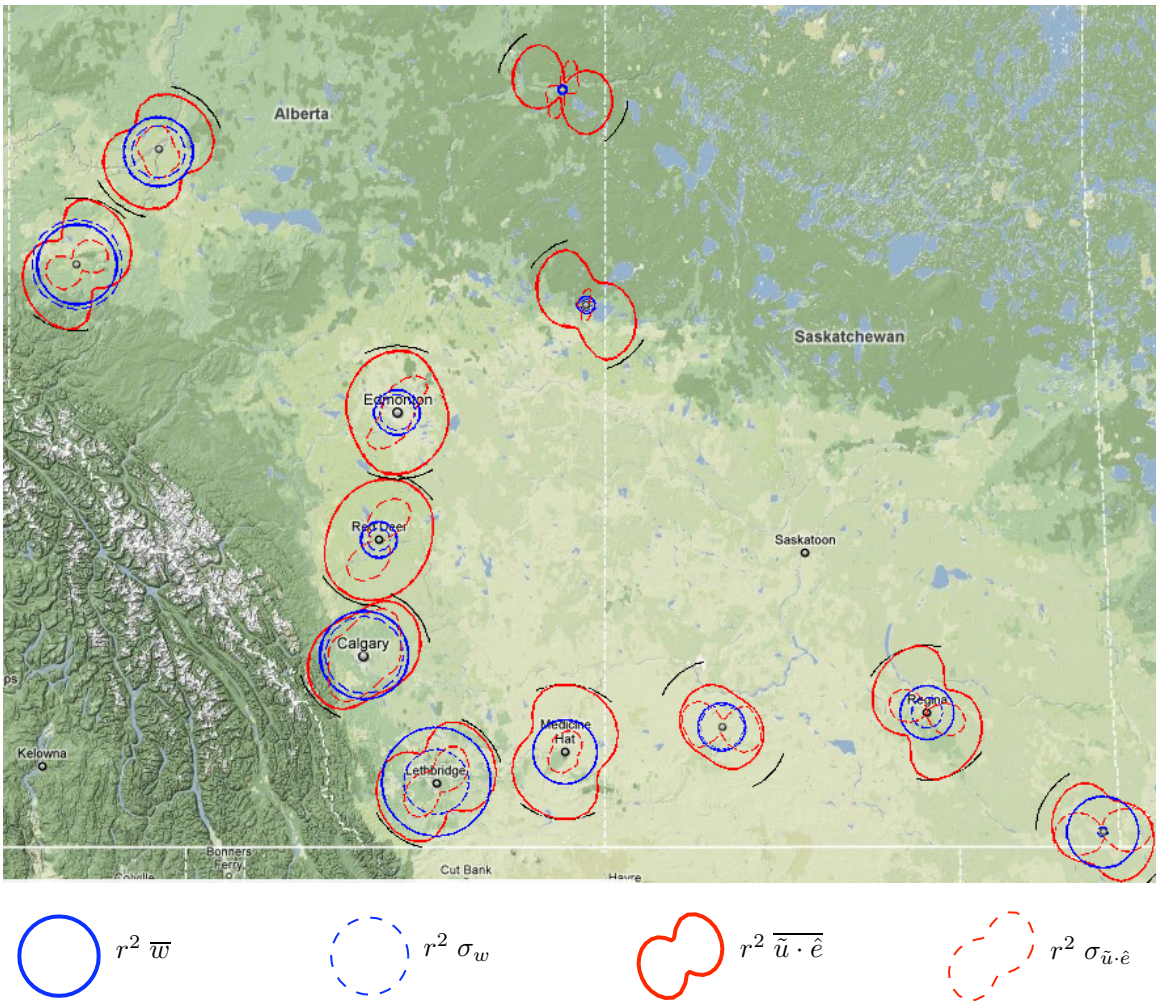

Figure 3.4: The variance explained (r^2) by predictions of the monthly statistics of the DJF winds observed at the Alberta and Saskatchewan sites. The black arcs represent a reference r^2 value of 0.6. Map obtained from Braxmeier (2011).

The primary differences in predictability across sites fall into three classes:

I/ **There are quantitative differences in the predictabiliy of the four predictands.** At each site there is some range of monthly-mean vector wind projections that are well predicted. In contrast, the predictability of the standard deviations and the mean wind speed vary considerably from site to site. There are sites where both the vector component predictands and the mean wind speed are well predicted (e.g. Lethbridge), while in other sites with nearly equivalent predictability of the vector predictands, the predictability of the mean wind speed is considerably smaller (e.g. Red Deer).

II/ **There is variability in the orientation of the anisotropy of vector wind predictability.** The orientation of the best predicted monthly-mean vector wind components varies across sites. There is also little consistency in the orientations of the anisotropy of the mean vector wind predictability and that of the vector wind standard deviation predictability. In many sites, these prediction skills share a similar orientation (e.g. Red Deer, Calgary, Lethbridge, and Medicine Hat), while in others the maximum predictability of the vector wind standard deviations falls along directions that yield a minimum predictability of the means (e.g. Peace River, Fort McMurray, and Cold Lake).

III/ **There are differences in the magnitude of the anisotropy of vector wind predictability.** In the case of Fort McMurray, the predictability of the mean vector component approaches zero at north-north-eastern (and south-south-western) projections, while at the other projections $r^2 \simeq 0.5$. On the other hand, the predictability of the mean components at sites such as Red Deer is much less sensitive to the direction of the vector projection.

Similar results were observed with the Ontario stations; these plots are included in Appendix C.

3.2.1 Systematic Results

Despite the large degree of differences in predictability characteristics between sites, there are some systematic results that hold across sites. These systematic results

are described below; an interpretation of these will be presented in Chapter 4. By first comparing the prediction skill of scalar predictands and vector predictands, it is observed that:

1. the means are better predicted than the standard deviations, particularly so with the vector predictands; and,
2. the vector wind predictands are better predicted than the scalar wind predictands.

Also, in agreement with previous results from van der Kamp et al. (2011) and Salameh et al. (2009) and discussed earlier,

3. there is a distinct anistropy to the prediction skill of the vector wind predictants.

As well as being true for the monthly wintertime predictions, these results are consistently found across season, averaging timescale, and location.

Kernel density estimates of the distribution of predictability (across site and season) of the scalar and vector winds are shown in Figure 3.5; to reduce the amount of information to be considered, we focus our attention on the best predicted vector components. The kernel density estimates shown in Figure 3.5 are used to provide a representation of the range of predictability of the surface wind quantities. The details of peaks in these estimates of the predictability distribution are sensitive to the width of the kernel used. A wide kernel (relative to the variability of the distribution) will smooth out features (i.e. peaks) in the distribution, while a narrow kernel will provide a much "noisier" estimate of the distribution of predictabilty. However, as these fine features will not interpreted in the current study, the sensitivity of the kernel density estimate to the kernel's width is not important for its representation of the range of predictive skills.

Figure 3.5 illustrates that in general the statistical downscaling (SD) model under consideration was quite poor at predicting the mean wind speed and the standard deviation of both scalar and vector winds. The SD model is particularly poor for predictions of seasonal wind speed quantities. On the other hand, the means of the

vector winds were generally reasonably well predicted, such that the best predictability occurred on monthly and daily averaging timescales. The predictability of the standard deviation of the vector wind demonstrates the opposite trend: the best predictions of sub-averaging timescale standard deviations tend to occur on longer averaging timescales. There is also a notable similarity between the distributions of the predictability of the scalar wind means and standard deviations on monthly and seasonal timescales.

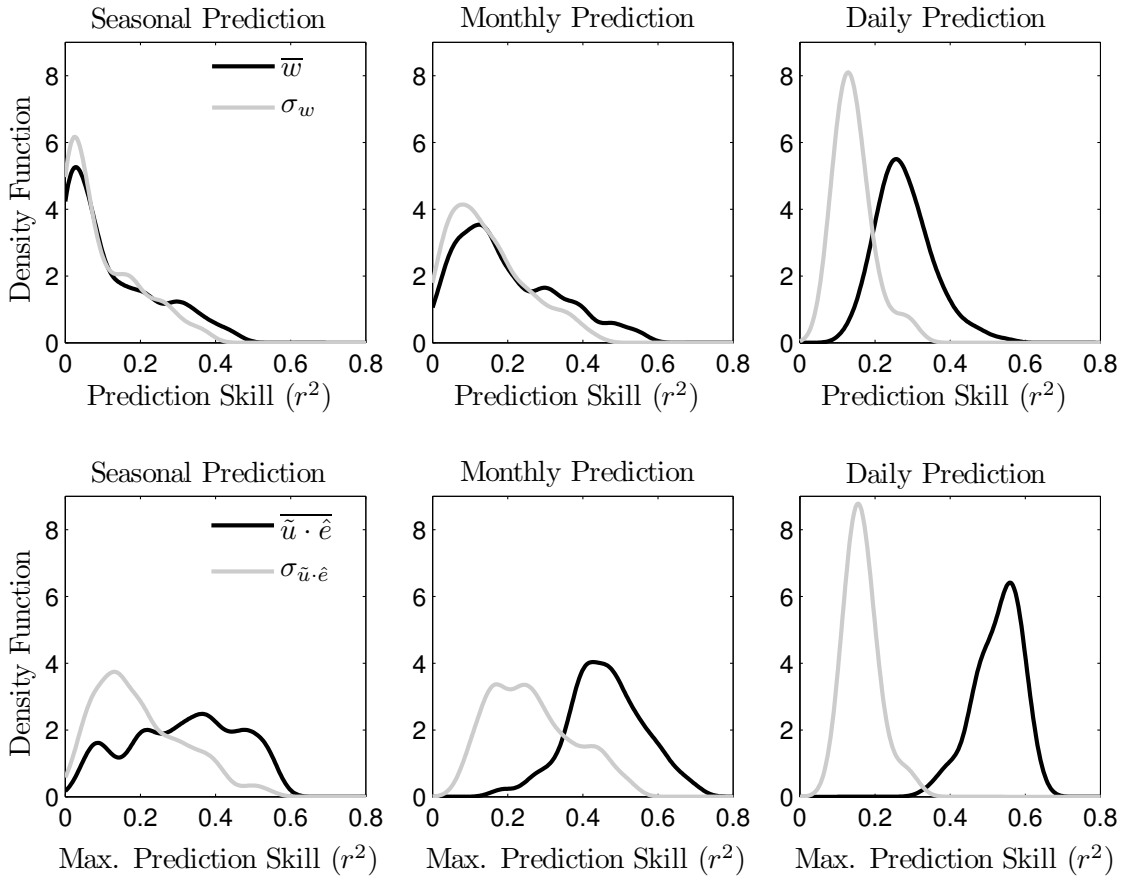

Figure 3.5: Kernel density estimates of the distributions of predictability (across site and season) of: the mean and standard deviation of the wind speed (top row); and the best predicted mean and standard deviation vector wind components (bottom row).

The mean and 75^{th} percentile of the predictability distributions shown in Figure 3.5 are below in Table 3.1. There does not appear to be a systematic dependence of the predictability of the standard deviations across averaging timescale. On the other hand, all means show improved predictability with shorter averaging timescales.

Table 3.1: Metrics of the cross-validated r^2 prediction skills (across site and season) for the scalar and the best predicted vector predictands. The values shown correspond to seasonal (ssn), monthly (mly), and daily (dly) averaging timescales.

Predictand	Mean r^2 (ssn — mly — dly)	75^{th} Percentile r^2 (ssn — mly — dly)
$\overline{w}$	0.12 — 0.20 — 0.28	0.21 — 0.30 — 0.32
σ_w	0.10 — 0.15 — 0.14	0.17 — 0.21 — 0.16
$\overline{\tilde{u} \cdot \hat{e}}$	0.31 — 0.46 — 0.53	0.43 — 0.53 — 0.57
$\sigma_{\tilde{u} \cdot \hat{e}}$	0.20 — 0.26 — 0.17	0.28 — 0.34 — 0.18

The strong similarities between some of the predictability distributions in Figure 3.5 are suggestive of relationships between prediction skills of different surface wind statistics. Scatterplots of the cross-validated r^2 prediction skill of the means and standard deviations for the monthly scalar (top row) and vector winds (bottom row) are shown in Figure 3.6. As with Figure 3.5 and Table 3.1, these scatterplots include results from all sites and seasons. There is evidence of a linear relationship between predictability of wind speed means and standard deviations, particularly on monthly timescales. When the mean wind speed is well predicted, so too is the standard deviation. On the other hand, there is no evident relationship between the predictability of the mean and standard deviation of the vector wind components on any averaging timescale.

We obtain three further systematic results across sites and seasons:

4. For a given set of predictor fields, a majority of the predictive information carried in the flow aloft relevant to the mean wind speed is also relevant to the standard deviation of wind speed.

5. For a given set of predictor fields, predictive information aloft relevant to the means of the wind vector projections is largely distinct from the information relevant to their standard deviations. A physical interpretation of this is that the degree of synoptic scale variability within a particular month or a season is not simply determined by the monthly- or seasonally-averaged state. Two months with similar mean free-tropospheric flows may display different degrees

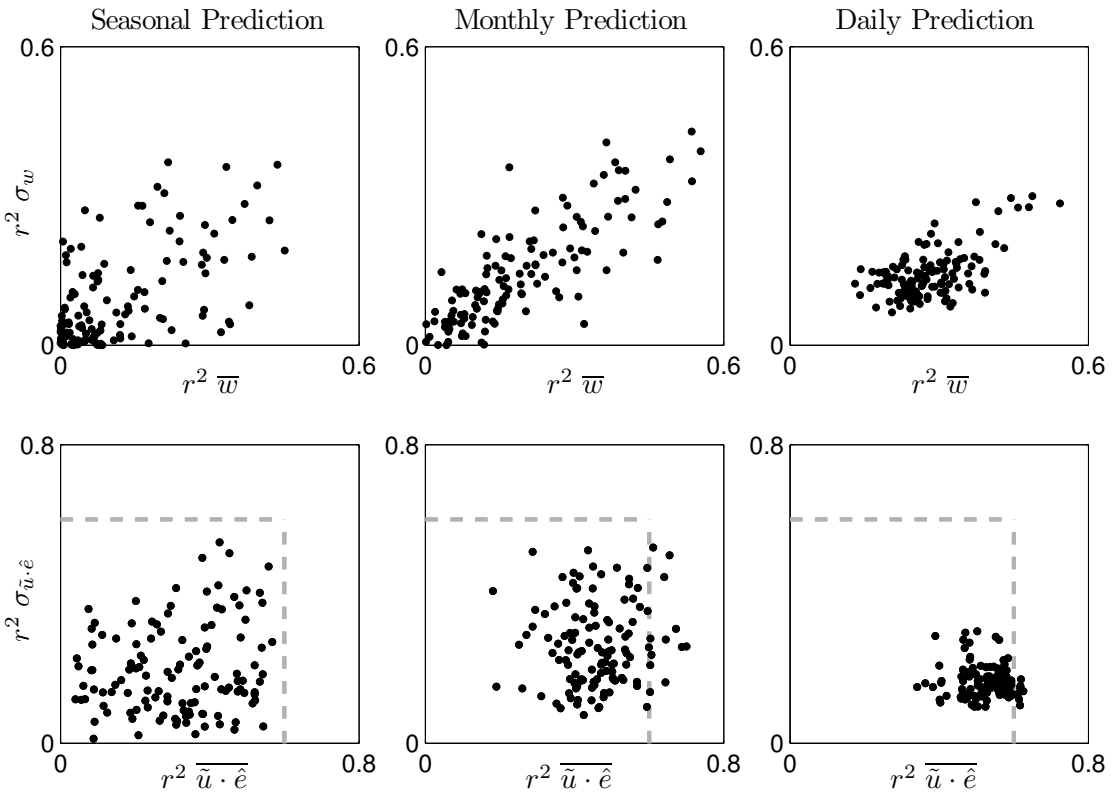

Figure 3.6: Scatterplots of cross-validated r^2 predictability of means and standard deviations for scalar (top row) and vector wind (bottom row) predictands. The vector wind scatterplots include only the best predicted component at each site. The dashed axes in the lower scatterplots correspond to the r^2 ranges of the scalar wind scatterplots (note the different axis scales between the upper and lower rows).

of storminess, and therefore different sub-monthly standard deviations.

6. Prediction skills of mean quantities are greater on shorter averaging timescales. In contrast, prediction skills of standard deviations do not display a systematic dependence on averaging timescale.

Systematic result 5 of the statistical predictions does not rule out the possibility of improving surface vector wind standard deviation predictions by using a different set of predictors aloft. Recall, however, that Chapter 2 reported the correlation of surface standard deviation time series from multiple stations and seasons with both the mean and the standard deviation of the predictors aloft. The corresponding correlation maps demonstrated weak correlations with small spatial scale structure.

On monthly and seasonal timescales, the mean winds are influenced by the low-frequency modes of the large-scale circulations. On weekly and daily timescales, the mean winds are influenced by the passage of cyclones and anticyclones (Robeson and Shein, 1997). As a result, the daily-averaged surface winds are much more representative of the dominant variability of the free troposphere, and hence better predicted than the monthly or seasonal-averaged surface winds.

3.2.2 Variations of predictability across season and site

We will now turn to assessing the sensitivity of surface wind predictability to

- the four calendar seasons,
- the location of the predictands.

The Seasonality of Predictability

Kernel density estimates of the distribution of predictability for each of the calendar seasons are displayed in Figure 3.7. There is no season where predictions clearly outperform all others. This result is in contrast with the study van der Kamp et al. (2011), which found predictability of monthly surface winds at sites in BC to be best in the wintertime and worst in the summertime. What weak seasonal variation of predictability is present is strongest on seasonal-mean timescales, and weakest for predictions on daily-mean timescales.

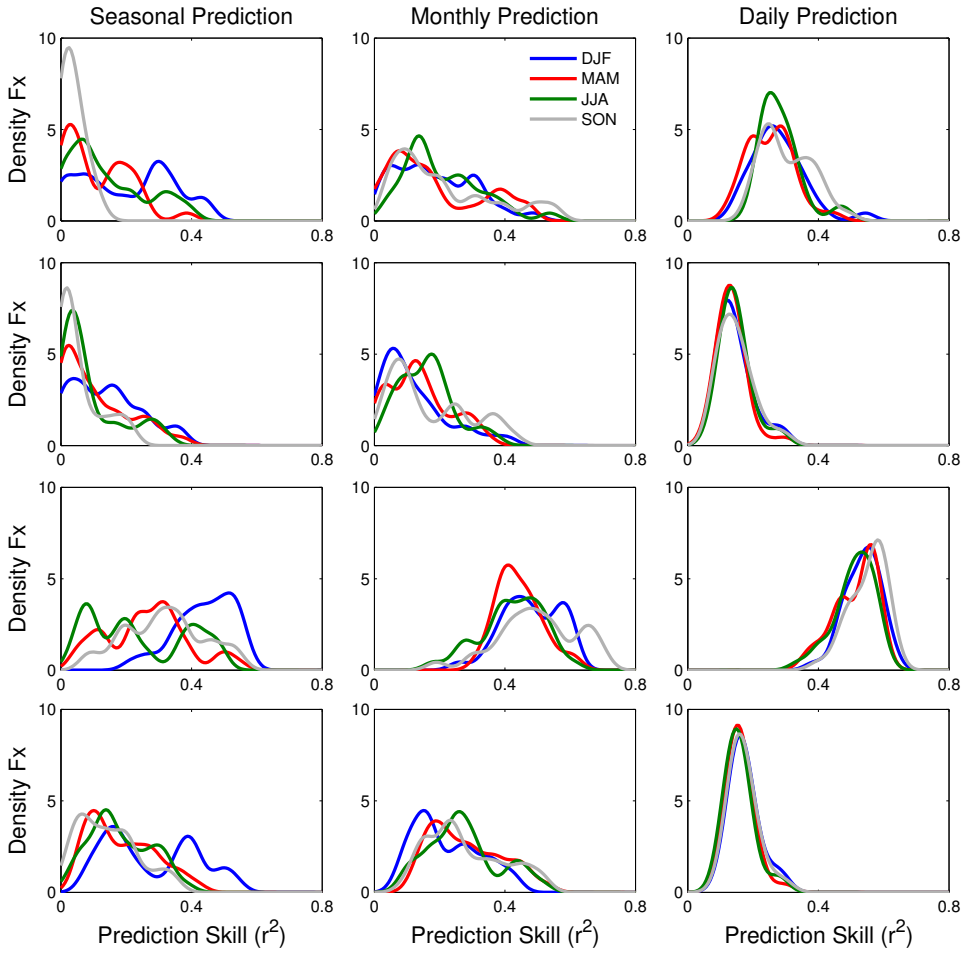

Figure 3.7: The seasonal dependence of the cross-validated r^2 prediction skill, stratified by predictand: $\overline{w}$ (top row), σ_w (second row), $\overline{\tilde{u} \cdot \hat{e}}$ (third row), and $\sigma_{\tilde{u} \cdot \hat{e}}$ (bottom row). Columns correspond to averaging timescale: seasonal (left column), monthly (middle column), and daily (right column). The season predicted is identified by the line colour; DJF (blue), MAM (red), JJA (green), SON (grey).

Station dependence of predictability

The preliminary analysis of relationships between surface wind variability at different sites (Chapter 1) suggested that the geography of the sites may have little impact on mean wind speed predictability. Figure 3.8 provides an overview of the geographic dependence of predictability of monthly-averaged DJF scalar wind quantities. Although the western sites have the highest predictability in this Figure, further plots in Appendix C demonstrate that there are no clear systematic differences in surface wind predictability between the three regions under consideration (Northern Plains, Southern Prairies, and Ontario).

The similar inter-region predictability is consistent with the horizontal extent and strength of the predictive information aloft associated with sites in these regions. The correlations of the monthly mean surface zonal wind at four locations with a suite of three large-scale predictor variables are shown in Figure 3.9. The white circle in each plot denotes the location of the surface station; these are Edmonton, AB; Estevan, SK; Kenora, ON; and Ottawa, ON. Only the Ottawa winds shows a noticeably different area of predictive information from the others, and this difference is limited to the mean zonal flow predictor. The location of the predictive information in the meridional and geopotential height anomaly fields is consistent across all four sites. The amount of information aloft relevant to the zonal means at Kenora is consistently greater than the others, suggestive of relatively weaker local influences. Consistent with this result, zonal winds at Kenora are the best predicted among these four sites.

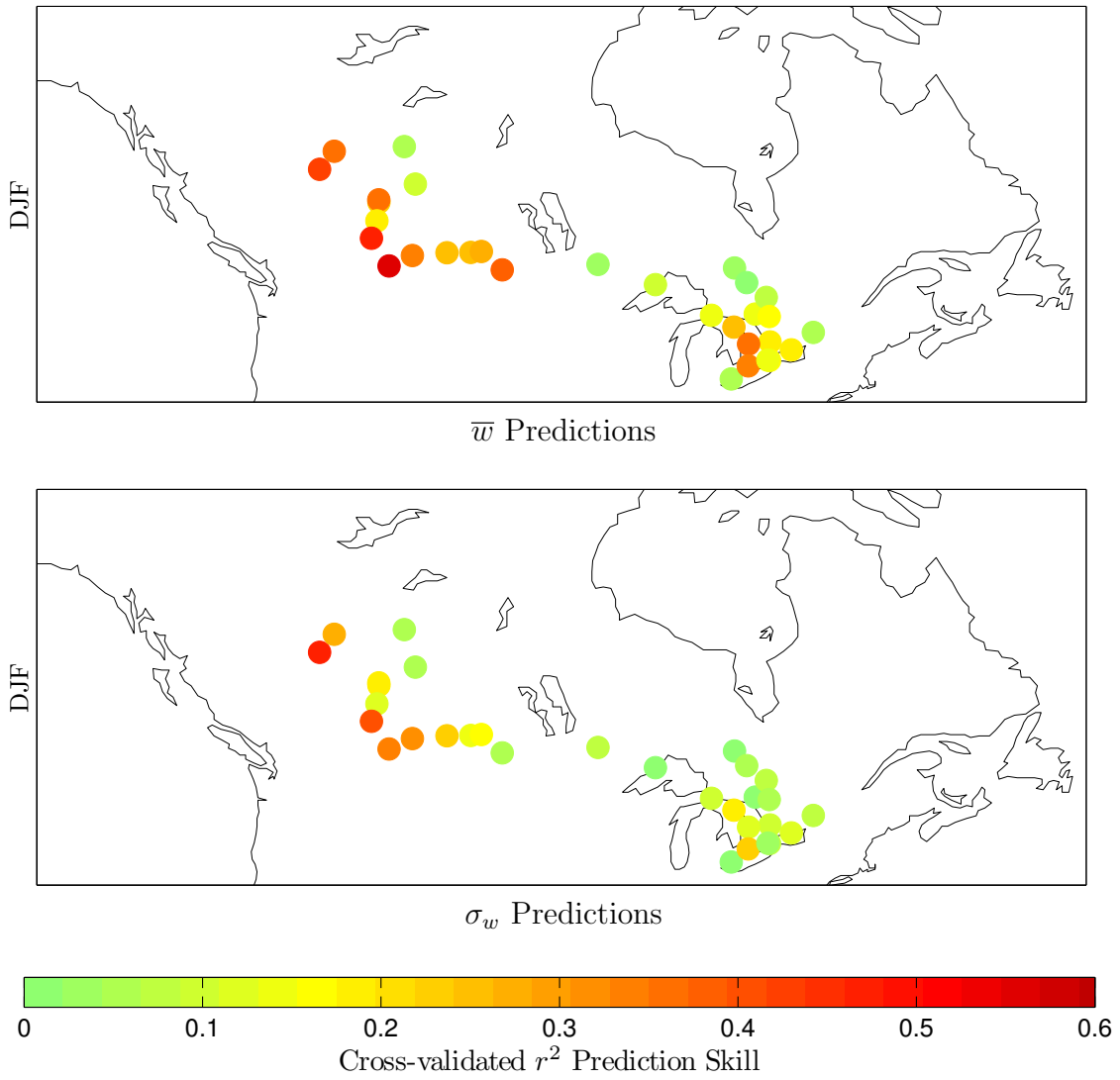

Figure 3.8: The cross-validated r^2 predictabilty of the DJF monthly scalar wind predictands: means, top row; and standard deviations, botton row. Plots for other seasons, predictands, and averaging timescales are presented in Appendix C.

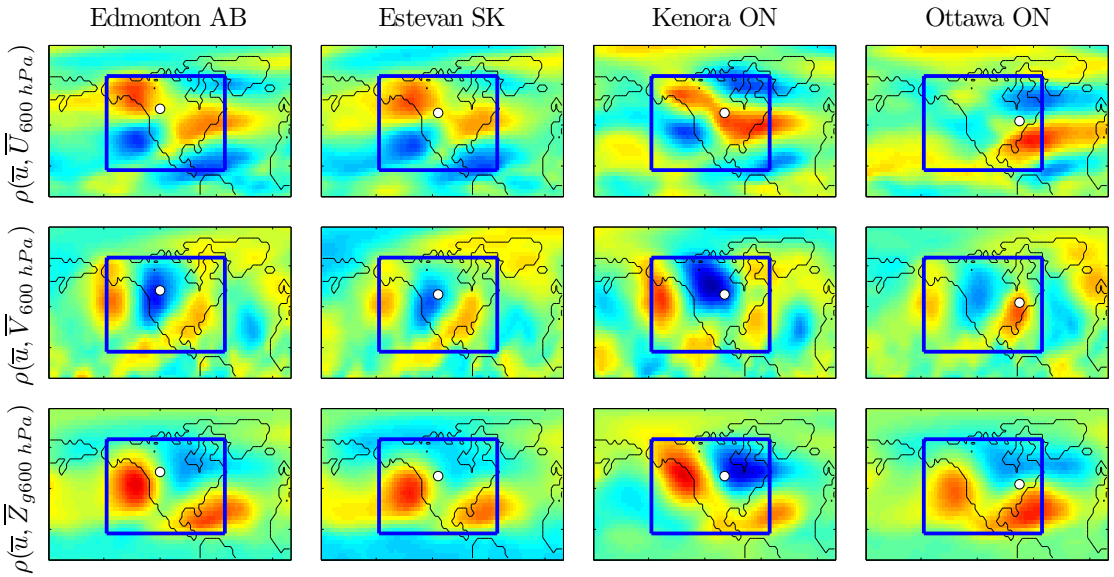

Figure 3.9: The correlation between the observed winter monthly mean surface zonal wind and free tropospheric variables at the 600 hPa NCEP/NCAR reanalysis pressure level. Each column contains correlation plots from a single observation site, the location of which is indicated by the white circle. Each row contains plots of the correlation between the observations and the monthly mean of one of three predictor variables: mean zonal flow (top row), mean meridional flow (middle row), and mean geopotential height (bottom row).

Chapter 4

Interpreting Predictability

The difference in the predictability of scalar and vector quantities is robust; it holds regardless of the season, the site, or the averaging timescale. This result may initially appear puzzling: after all, the wind speed is simply the magnitude of the vector winds:

$$w^2 = u^2 + v^2. \tag{4.1}$$

The following discussion will seek to demonstrate that the disparity in predictability is in fact readily explainable.

First, we note that the equality in Eqn. 4.1 is only valid at a particular instant in time; once these quantities are averaged, the equality stops:

$$\overline{w}^2 \neq \overline{u}^2 + \overline{v}^2. \tag{4.2}$$

Consider a hypothetical site located along the coast of a large body of water, where the direction of the surface temperature gradient reverses from day to night. As the land temperature rises in the daytime an on-shore flow is generated. At night, when the land cools and its temperature drops below that of the body of water, an offshore wind results. Averaged over a daily time scale, if the daytime onshore winds are equally strong as the offshore nighttime winds, the mean wind speed will be non-zero while the mean vector wind will be zero. The mean wind speed is not a function of the mean vector wind components alone.

4.1 The Idealized Probability Distribution Model

In Monahan (2012a), an idealized model (IPM) of the wind speed probability distribution function (PDF) was introduced to study the mean wind speed as a function of the vector wind statistics. The IPM assumes a bivariate Gaussian distribution of uncorrelated vector wind components:

$$p_{uv}(u,v) = \frac{1}{2\pi\sigma_u\sigma_v}\exp\left(-\frac{(u-\overline{u})^2}{2\sigma_u^2}-\frac{(v-\overline{v})^2}{2\sigma_v^2}\right), \tag{4.3}$$

where

$\overline{u}$, $\overline{v}$ are the means of orthogonal vector winds, and
σ_u, σ_v are the standard deviations of orthogonal vector winds.

With the polar coordinate transformation

$$u = w\cos\phi, \tag{4.4}$$
$$v = w\sin\phi, \tag{4.5}$$

such that

$$w\,dw\,d\phi = du\,dv, \tag{4.6}$$

the probability distribution of u and v may be transformed to a conditional PDF of wind speed and direction. The marginal PDF of the wind speed may then be obtained by integrating the conditional PDF over the wind's direction, ϕ,

$$p(w)dw = \frac{w}{2\pi\sigma_u\sigma_v}\int_{-\pi}^{\pi}\exp\left(-\frac{(w\cos\phi-\overline{u})^2}{2\sigma_u^2}-\frac{(w\sin\phi-\overline{v})^2}{2\sigma_v^2}\right)d\phi\,dw. \tag{4.7}$$

An expression for the mean wind speed ($\overline{w}$) is found by integrating the marginal PDF of the wind speed,

$$\overline{w} = \int_0^{\infty} wp(w)dw. \tag{4.8}$$

The result is an expression relating the mean wind speed to four statistical features of the vector winds: $\overline{u}$, $\overline{v}$, σ_u, and σ_v.

The expression for $\overline{w}$ can be simplified if it is assumed that the vector wind fluctuations are isotropic: $\sigma_u = \sigma_v = \sigma$. It can then be shown that

$$\overline{w} = \sigma\sqrt{\frac{\pi}{2}}\exp\left(\frac{-\mu}{2\sigma}\right)^2\left[\left(1+2\left(\frac{\mu}{2\sigma}\right)^2\right)I_0\left(\frac{\mu}{2\sigma}\right)^2+2\left(\frac{\mu}{2\sigma}\right)^2 I_1\left(\frac{\mu}{2\sigma}\right)^2\right], \quad (4.9)$$

where $I_k(z)$ is the associated Bessel Function of the first kind of order k (Monahan, 2012b) and

$$\mu = \sqrt{\overline{u}^2+\overline{v}^2} \quad (4.10)$$

is the magnitude of the mean vector wind. The two-variable IPM efficiently distributes the information relevant to the wind speed across two characteristics of the vector winds. Note that the expression in Eqn. 4.9 is of the form

$$\overline{w} = \sigma F\left(\frac{\mu}{\sigma}\right). \quad (4.11)$$

A similar two-variable expression may be found for the standard deviation of the wind speed, σ_w;

$$\begin{aligned}\sigma_w^2 = \mathrm{E}[(w-\overline{w})^2] &= \mathrm{E}[w^2]-\overline{w}^2 \\ &= \sigma_u^2+\overline{u}^2+\sigma_v^2+\overline{v}^2-\overline{w}^2 \quad (4.12)\\ &= 2\sigma^2+\mu^2-\overline{w}^2 \quad (4.13)\end{aligned}$$

where E[.] denotes mathematical expectation. It follows that the wind speed standard deviation is given by

$$\sigma_w = \sigma\sqrt{2+\left(\frac{\mu}{\sigma}\right)^2-F\left(\frac{\mu}{\sigma}\right)^2}. \quad (4.14)$$

4.1.1 Robustness of the IPM

Before the IPM is used to interpret our predictability results, it is important to investigate the robustness of the assumptions that it is based upon. The IPM was first introduced for the analysis of sea surface winds over the sub-Arctic Pacific Ocean

(Monahan, 2012a). The assumptions it is based upon may not be as reasonable for winds over land. These assumptions are

i. a bivariate Gaussian distribution;

ii. uncorrelated mean vector winds; and,

iii. the approximation of isotropic vector wind variability.

To quantify the error introduced by these approximations, the monthly mean wind speeds (for each site and season) were calculated from monthly mean vector wind statistics with Eqn. 4.9 ($\overline{w}_{IPM}$) and compared to monthly mean wind speeds computed directly from observations ($\overline{w}_{Obs}$; Figure 4.1, left plot). The isotropic variance σ^2 was estimated as the mean of the variances of the vector wind components:

$$\sigma^2 = \frac{1}{2}(\sigma_u^2 + \sigma_v^2). \tag{4.15}$$

The agreement between the quantities calculated with the IPM and those computed directly from observations is generally excellent. The ability of the IPM to accurately reproduce mean wind speed variability at each site and season was also evaluated (Figure 4.1, right plot). Aside from Earlton ON, the direct correlation of the two quantities is everywhere greater than 0.97. Despite its approximations, the idealized model provides a good characterisation of the mean wind speed.

4.1.2 Wind Speed Sensitivities

The two-variable IPM has one particularly strong advantage over the more general four-variable version: the sensitivities of the statistical features of the wind speed ($\partial_\mu \overline{w}$, $\partial_\sigma \overline{w}$, $\partial_\mu \sigma_w$, and $\partial_\mu \sigma_w$,) are found to be functions of the dimensionless ratio $\frac{\mu}{\sigma}$. From Eqn 4.11 it follows that the sensitivities $\partial_\mu \overline{w}$ and $\partial_\sigma \overline{w}$ are:

$$\partial_\mu \overline{w} = F'\left(\frac{\mu}{\sigma}\right) \tag{4.16}$$

$$\partial_\sigma \overline{w} = F\left(\frac{\mu}{\sigma}\right) - \frac{\mu}{\sigma} F'\left(\frac{\mu}{\sigma}\right) \tag{4.17}$$

$$= F\left(\frac{\mu}{\sigma}\right) - \frac{\mu}{\sigma}\partial_\mu \overline{w}. \tag{4.18}$$

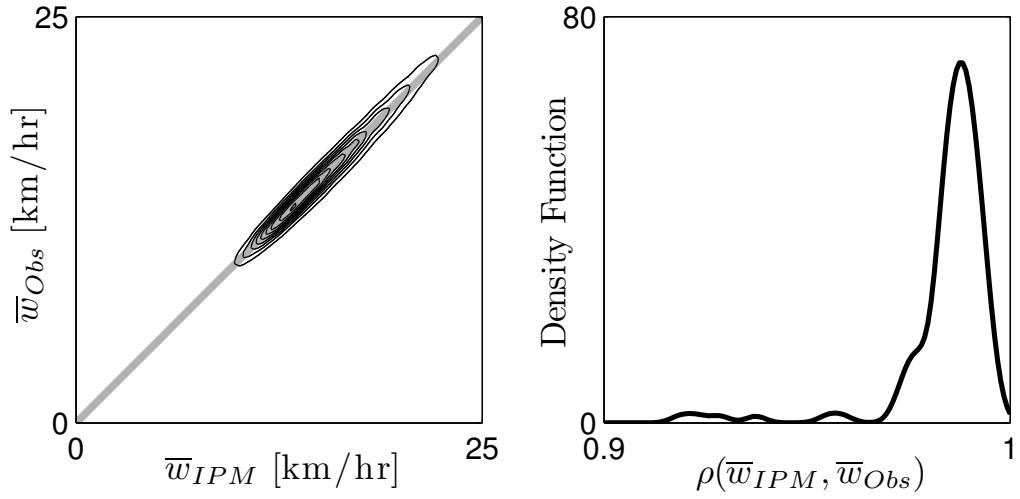

Figure 4.1: Kernel density estimates of the joint probability distribution (across site and season) of monthly $\overline{w}$ computed directly from observations ($\overline{w}_{Obs}$) and values calculated with mean vector wind statistics and Eqn. 4.9 ($\overline{w}_{IPM}$) (left). The right panel displays the probability distribution of direct correlations of $\overline{w}_{Obs}$ and $\overline{w}_{IPM}$.

Expressions for the sensitivities $\partial_\mu \sigma_w$, and $\partial_\sigma \sigma_w$ are somewhat more complex:

$$\partial_\mu \sigma_w = \frac{\mu - \sigma F\left(\frac{\mu}{\sigma}\right) F'\left(\frac{\mu}{\sigma}\right)}{\sigma_w} \tag{4.19}$$

$$= \frac{\mu - \sigma F\left(\frac{\mu}{\sigma}\right) \partial_\mu \overline{w}}{\sigma_w}, \tag{4.20}$$

$$\partial_\sigma \sigma_w = \frac{2\sigma - \sigma F^2\left(\frac{\mu}{\sigma}\right) + \mu F\left(\frac{\mu}{\sigma}\right) F'\left(\frac{\mu}{\sigma}\right)}{\sigma_w} \tag{4.21}$$

$$= \frac{\sigma[2 - F^2\left(\frac{\mu}{\sigma}\right)] + \mu F\left(\frac{\mu}{\sigma}\right) \partial_\mu \overline{w}}{\sigma_w}. \tag{4.22}$$

Because the sensitivities depend on the ratio $\dfrac{\mu}{\sigma}$ alone, not these parameters separately, it is convenient to introduce the variable

$$\theta = \arctan\left(\frac{\mu}{\sigma}\right). \tag{4.23}$$

The sensitivities $\partial_\mu \overline{w}$, $\partial_\sigma \overline{w}$, $\partial_\mu \sigma_w$, and $\partial_\sigma \sigma_w$ as functions of θ are displayed in Figure 4.2. There are several points of key significance. First, the end members of the θ range represent two distinct regimes of mean wind speed sensitivities. In a low θ regime,

the variability of $\overline{w}$ is dominated by the variability of σ; essentially all the information relevant to the variability of mean wind speeds is held in σ. At the other θ extreme, the variability of $\overline{w}$ is dominated by that of μ. Second, μ exerts very little influence over wind speed standard deviation; the variability in σ_w is determined primarily by the vector winds' standard deviation (σ).

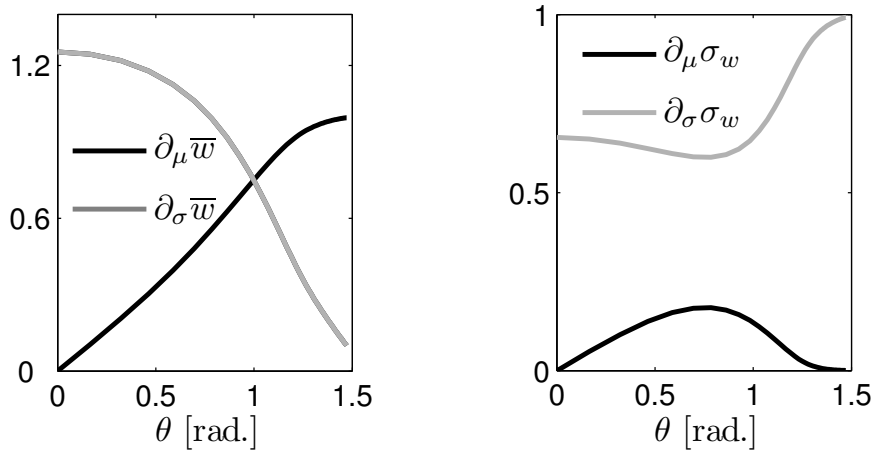

Figure 4.2: The sensitivities of the wind speed statistics to μ (the magnitude of the mean vector winds) and to σ (the isotropic fluctuations of the vector winds) as a function of $\theta = \arctan(\mu/\sigma)$.

For the hypothetical waterfront site (described at the start of this chapter), the daily θ would be zero (since the mean vector wind is zero), thus changes in the mean wind speed would be dictated solely by the vector winds' variability, σ. Conversely, at sites characterised by largely steady winds from a particular direction, μ is expected to be much greater than σ and mean wind speeds will respond to changes in μ alone. In both of these sites, if θ is sufficiently small or large, μ will hold essentially no influence over the standard deviation of the wind speed.

We will use the IPM characterisations of the sensitivities of $\overline{w}$ and σ_w to the statistical features of the vector winds to interpret the observed differences in predictability between scalar and vector wind statistics.

4.2 The predictability of $\overline{w}$ relative to μ and σ

In order to use the wind speed sensitivities to explain wind speed predictability, it is necessary to establish the θ ranges of the surface winds under consideration. Histograms of the mean θ values (on seasonal, monthly, and daily timescales) observed at the sites under consideration are displayed in Figure 4.3. It is clear that longer averaging timescales are associated with lower θ values. Within this θ range, the IPM predicts the information relevant to mean wind speed's variability will be carried primarily by σ. At daily timescales, the observations are found to be in an intermediate θ within which the information relevant to $\overline{w}$ is predicted to be partitioned approximately equally between μ and σ.

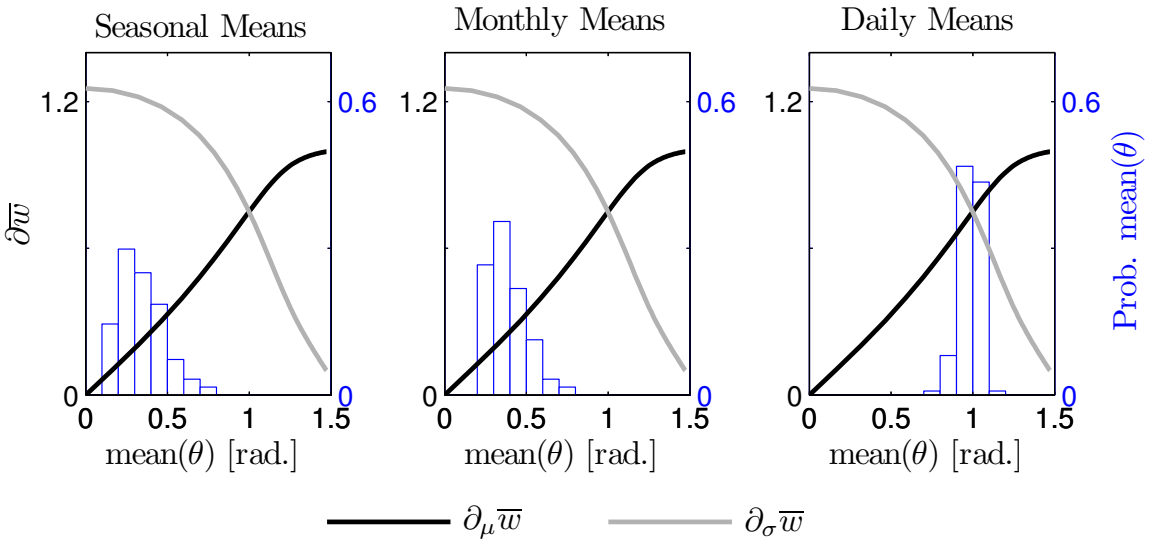

Figure 4.3: The frequency probability (right axis) of the observed winds' mean θ value on seasonal (left), monthly (middle), and daily (right) averaging timescales. The sensitivities of the mean wind speed to μ and σ are overlayed (black and grey curves, respectively).

To evaluate the utility of the IPM in interpreting the predictability of $\overline{w}$, predictions of μ and σ were made and compared to corresponding predictions of mean wind speed. Note that predicting μ directly is not in general the same thing as predicting $\overline{u}$ and $\overline{v}$ separately and then combining these into a prediction of μ, as these quantities are nonlinearly related (Monahan, 2012b). This point is discussed in more detail in Section 4.3. Scatterplots of the cross-validated r^2 prediction skill of μ and σ compared

to $\overline{w}$ predictability (for all seasons and locations) are displayed in Figure 4.4 for each of the averaging timescales considered. There is a strong linear relationship between the predictability of $\overline{w}$ and of σ on both seasonal and monthly timescales. On both averaging timescales, the correlations (ρ) between the r^2 values of $\overline{w}$ and of σ is 0.87 ($\rho^2 = 0.75$). This is in stark contrast to the weak correlations between the r^2 values of $\overline{w}$ and of μ; here ρ is 0.44 ($\rho^2 = 0.19$) on seasonal timescales, and 0.36 ($\rho^2 = 0.13$) on monthly timescales. This is consistent with the expectations from the IPM (e.g. Figure 4.3).

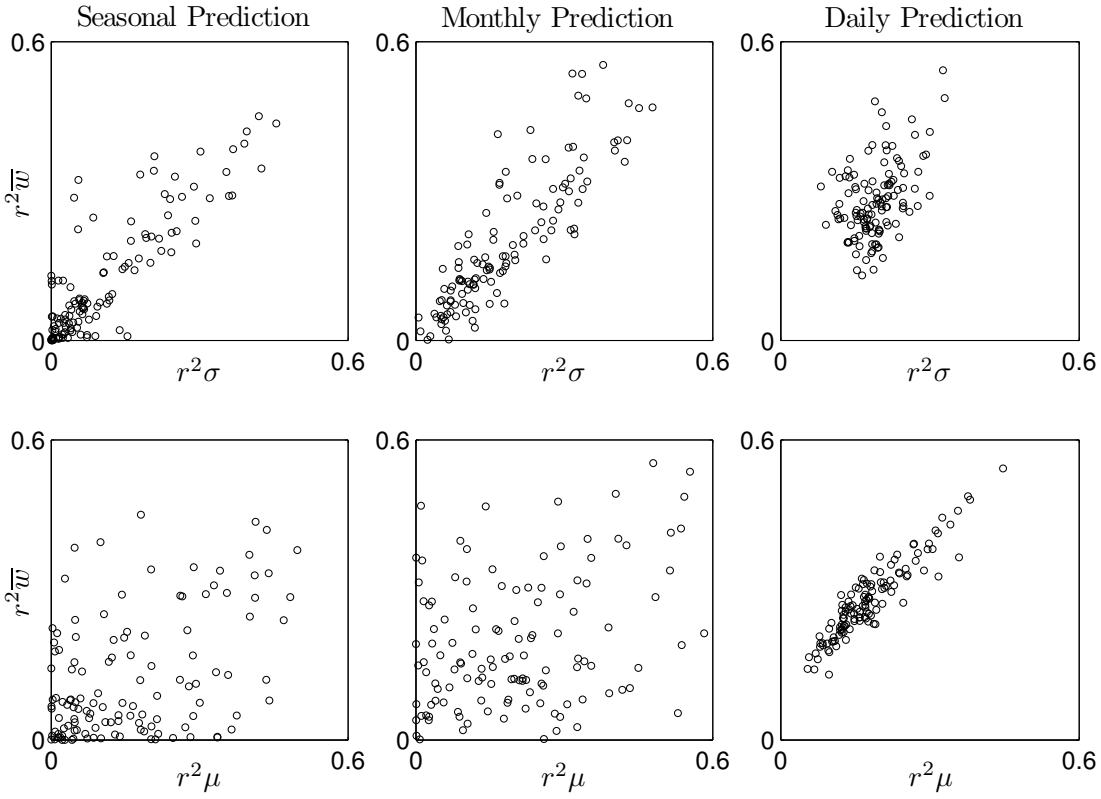

Figure 4.4: Scatterplots of the cross-validated r^2 prediction skills of $\overline{w}$ against those of μ and σ on seasonal, monthly, and daily averaging timescales.

On daily timescales, there are linear relationships between the predictability of $\overline{w}$ and both μ and σ. The direct correlation of the predictability of $\overline{w}$ and of σ is 0.44 ($\rho^2 = 0.20$), while a high correlation of 0.93 ($\rho^2 = 0.86$) is found between the predictability of $\overline{w}$ and of μ. Since the IPM predicts an approximately equal sensitivity of $\overline{w}$ to both variables, the contrasting strengths of these correlations is somewhat surprising at first. In fact, the relative influences of μ and σ on $\overline{w}$ will be

determined both by the sensitivities, $\partial_\mu \overline{w}$ and $\partial_\sigma \overline{w}$, and by the absolute variability, std(μ) and std(σ).

To demonstrate this, a scatterplot of the long-term standard deviations of the daily std(μ) and std(σ) timescale quantities (for each location and season) is displayed in Figure 4.5. The standard deviations of μ are generally two to three times greater in size than the standard deviations of σ. The relative sizes of the absolute variabilty of the μ and σ explains the greater influence on daily timescales of μ on $\overline{w}$ than σ despite the approximately equal sensitivities.

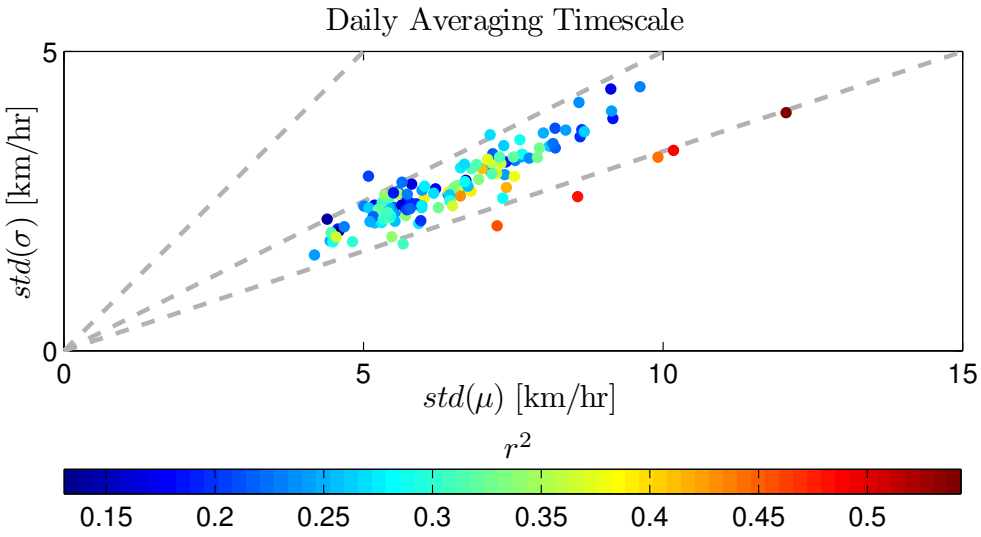

Figure 4.5: The long-term standard deviations of daily μ and σ (across site and season); the marker colour denotes the cross-validated r^2 skill of the corresponding $\overline{w}$ predictions. The dashed lines represent std(σ):std(μ) ratios of 1:1, 1:2, and 1:3.

The colours of the points in Figure 4.5 indicate the predictability of $\overline{w}$ found at that particular site and season. There is no evident relationship between the day-to-day variability of either μ or σ individually and the predictability of $\overline{w}$. However, there is evidence of a weak dependence of mean wind speed predictability on the relative absolute variability of μ and σ. That is, the predictions of $\overline{w}$ tend to be more skillful as the magnitude of the vector means (μ) becomes much more variable than the isotropic standard deviation of the vector winds (σ).

4.3 The predictability of μ relative to $\overline{u}$ and $\overline{v}$

In the previous section, the IPM was found to provide a useful explanation of the relative predictability of the vector and scalar wind variability on seasonal, monthly, and daily timescales. The dependence of $\overline{w}$ on σ in low θ regimes, and on μ in high θ regimes, leads to the question: are wind speeds in a high θ regime inherently more predictable than those in a low θ regime? If the answer is yes, a site's θ value will provide an a priori qualification of wind speed predictability. In Chapter 3, it was shown that the vector wind means $\overline{\tilde{u} \cdot \hat{e}}$ are generally better predicted than their standard deviations $\sigma_{\tilde{u} \cdot \hat{e}}$. Does it follow from this that μ is better predicted than σ?

The predictability of μ and σ may be assessed by revisiting the scatterplots displayed in Figure 4.4. While μ and σ display a wide range of predictive skills, on average these skills are not particularly large. A further illustration of this fact is seen in the r^2 probability distributions (across season and site) of seasonal, monthly, and daily quantities of μ and σ displayed in Figure 4.6 (top row). For the locations under consideration, there is no substantial difference in the predictive skills of μ and σ. In contrast, the means of the vector wind components are much better predicted on all timescales (Figure 4.6, bottom row).

The poor predictability of μ is seemingly at odds with the relatively good predictability of the mean vector wind components. If μ were a linear function of the vector wind means, a strong relation between its predictability and those of the mean vector wind components would be expected; but μ is a nonlinear transformation of the mean vector winds. It is in part because of this non-linearity that there will be instances where both of the orthogonal mean vector winds may be quite well predicted while μ will have no linear predictability.

Consider the monthly zonal and meridional winds at Edmonton during the autumn months from 1961 to 2006 displayed in Figure 4.7. The zonal and meridional monthly-mean wind components are shown, as well as the corresponding value of μ. An examination of the linear relationship between μ and the monthly mean vector wind components sheds light on the influence of the non-linear tranformation. Over this period, 29% of the variance of μ is linearly explained by $\overline{u}$, while 6% is explained by $\overline{v}$. Although there is some shared linear information, it is small. This example

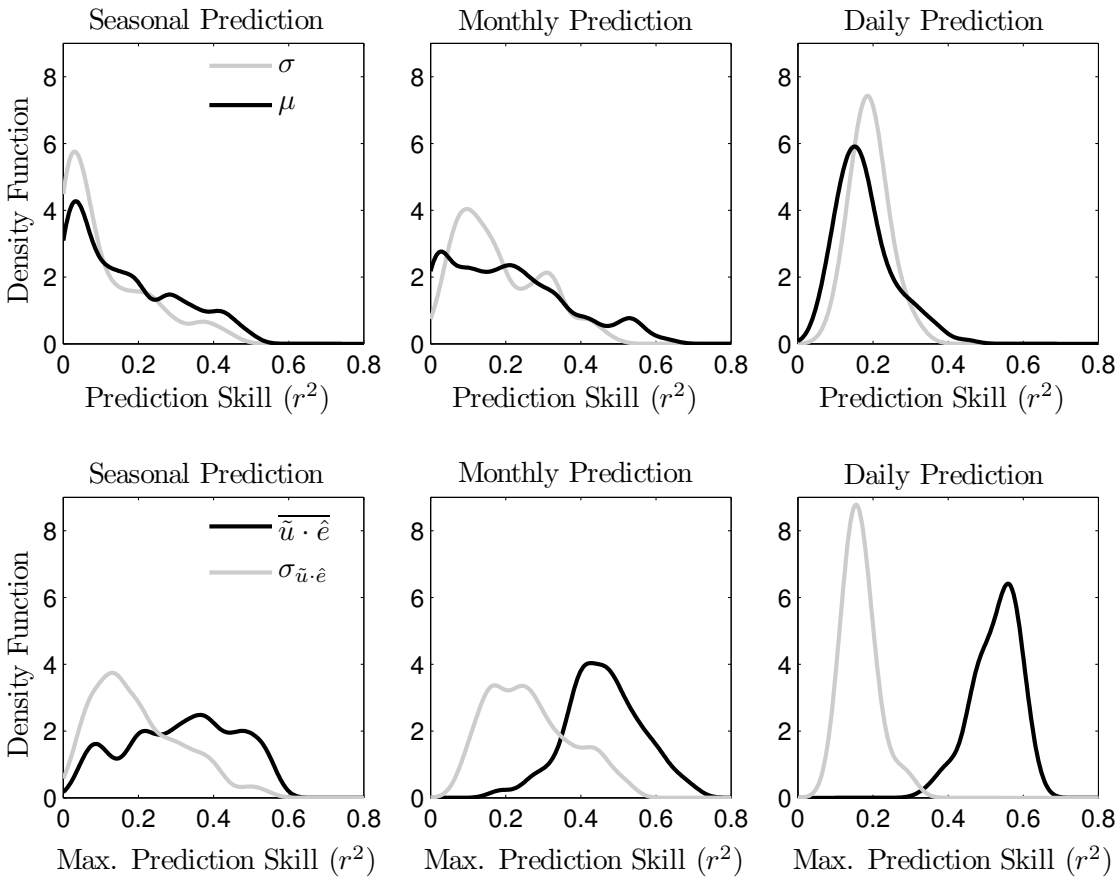

Figure 4.6: Kernel density estimates of the distributions of predictability (across site and season) of μ and σ (top row); and the best predicted mean and standard deviation vector components (bottom row).

clearly shows there can be weak correlation between μ and one or both of the orthogonal vector wind means, so good predictability of $\overline{u}$ and $\overline{v}$ may not imply good predictability of μ.

A discussion of the predictabilty of $\mu^2 = \overline{u}^2 + \overline{v}^2$ in terms of the predictability of $\overline{u}$ and of $\overline{v}$ is presented in Monahan (2012b). This earlier study takes advantage of the fact that to an excellent approximation, for a predictor x,

$$cov(x, \mu^2) \simeq cov(x, \mu) \tag{4.24}$$

where $cov(x, \mu)$ is the covariance of x and μ. Equivalently, the approximation that corr(μ,μ^2) $\simeq$ 1 is made. Probability distributions of corr(μ, μ^2) (across site and

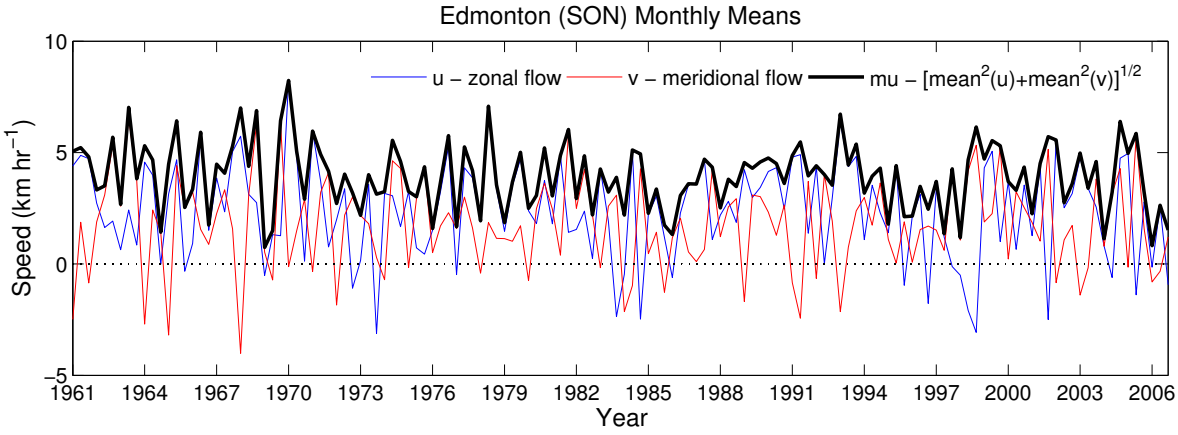

Figure 4.7: The observed monthly mean zonal wind, meridonal wind, and vector wind amplitude at Edmonton during the autumn months of 1961 to 2006.

season) at the averaging timescales considered are displayed in Figure 4.8. The correlations are always strong, although weaker at shorter averaging timescales.

To justify the approximation in Eqn 4.24, Monahan (2012b) used the following heuristic approach. By definition:

$$\mu^2 = (mean(\mu) + \mu' std(\mu))^2 \tag{4.25}$$

$$= std^2(\mu)\left[\frac{mean^2(\mu)}{std^2(\mu)} + 2\frac{mean(\mu)}{std(\mu)}\mu' + \mu'^2\right], \tag{4.26}$$

where μ' is the standardized anomaly with mean zero and unit variance:

$$\mu' = \frac{\mu - mean(\mu)}{std(\mu)}. \tag{4.27}$$

When the ratio of mean(μ) over std(μ) is greater than one, the linear dependence of μ^2 on fluctuations of μ' dominate and to a first order approximation Eqn. 4.24 is satisfied. Figure 4.9 illustrates the distributions (across site and season) of this ratio at the three averaging timescales under consideration. That this ratio is everywhere greater than one and lowest on the daily averaging timescale is consistent with the empirically calculated values of $corr(\mu, \mu^2)$ displayed in Figure 4.8.

With this approximation, expressions for the predictability of μ in terms of the predictability of $\overline{u}$ and $\overline{v}$ can be obtained if it is assumed that fluctuations in $\overline{u}$ and

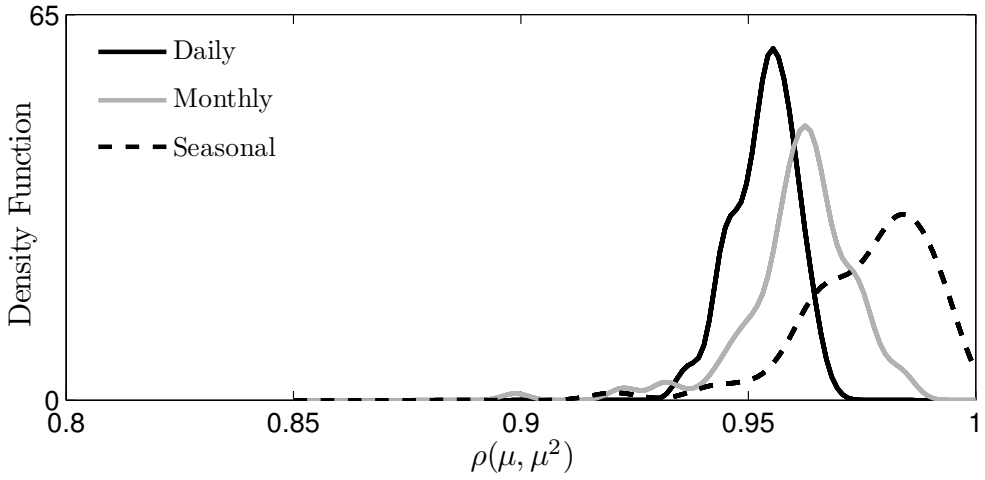

Figure 4.8: Kernel density estimates of the probability distributions of the direct correlation of μ and μ^2 (across site and season) observed on seasonal, monthly, and daily averaging timescales.

$\overline{v}$ are:

i. uncorrelated, and

ii. isotropic [$\kappa = std(\overline{u}) = std(\overline{v})$].

Denoting the respective predictions of the mean vector winds $\overline{u}$ and $\overline{v}$ as

$$corr(x, \overline{u}) = \rho_{\overline{u}} \tag{4.28}$$

$$corr(x, \overline{v}) = \rho_{\overline{v}}, \tag{4.29}$$

the linear predictability of μ is expressed as:

$$corr(x, \mu) \simeq \frac{mean(\overline{u})\rho_{\overline{u}} + mean(\overline{v})\rho_{\overline{v}}}{\sqrt{\kappa^2 + mean^2(\overline{u}) + mean^2(\overline{v})}}. \tag{4.30}$$

We can align the coordinate system such that $\overline{u}$ and $\overline{v}$ are the along-mean and cross-mean winds. In this case $mean(\overline{v}) = 0$, and so

$$corr(x, \mu) \simeq \frac{mean(\overline{u})\rho_{\overline{u}}}{\sqrt{\kappa^2 + mean^2(\overline{u})}}. \tag{4.31}$$

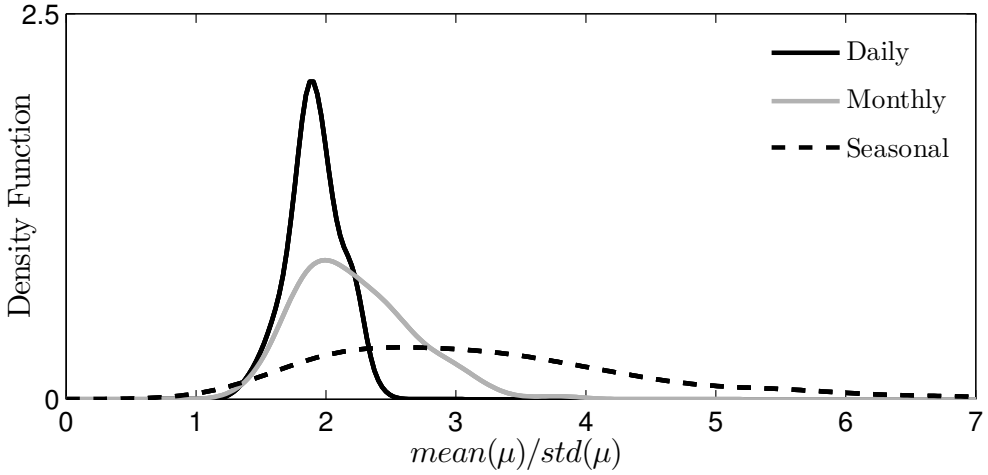

Figure 4.9: Kernel density estimates of the probability distributions of the ratio of mean(μ) over std(μ) (across site and season) observed on seasonal, monthly, and daily timescales.

At sites where the prevailing winds dominate, $mean(\overline{u}) \gg \kappa$ and

$$corr(x, \mu) \simeq \rho_{\overline{u}}. \tag{4.32}$$

However, where $\kappa >> mean(\overline{u})$, $corr(x, \mu) << \rho_{\overline{u}}$. Probability distributions of the ratios

$$\frac{\kappa}{mean(\overline{u})} \quad \text{and} \quad A^2, \text{ where } A = \frac{mean(\overline{u})}{\sqrt{\kappa^2 + mean^2(\overline{u})}}$$

are displayed in Figure 4.10. In both plots, there is a strong similarity between the distributions of seasonally- and monthly-averaged quantities. On these averaging timescales, we find that κ is generally smaller than $mean(\overline{u})$ and the predictability of μ is close to that of $\overline{u}$. We also find that decreases in the averaging timescales, κ increases and (hence) the predictability of μ relative to the mean vector winds decreases.

The sensitivity of the predictability of μ to the averaging timescale is a result of the averaging procedure partitioning surface wind variability into resolved (means) and unresolved (standard deviations) components. The long-term mean, $mean(\overline{u})$, is independent of averaging timescale; however, the variability of the vector mean (κ) varies with averaging timescales. In general, as the averaging timescales increases,

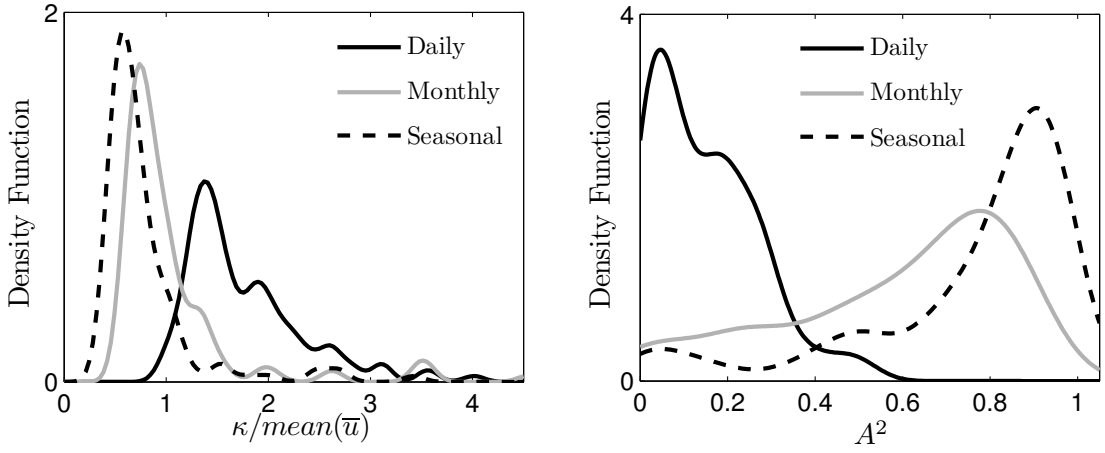

Figure 4.10: Kernel density estimates of the probability distributions of the ratio of: (κ) over $mean(\overline{u})$, at left; and the ratio A^2, at right, (across site and seasons) observed on seasonal, monthly, and daily timescales.

more of the wind variability is on unresolved sub-averaging timescales so κ decreases and $corr(x, \mu)$ approaches $corr(x, \overline{u})$ (equivalently A approaches 1; Figure 4.10, right plot).

Since θ compares resolved and unresolved vector wind variability, it shares a similar sensitivity to the averaging timescale. The variable θ will tend to be small where the variability is on sub-averaging timescales; otherwise it is large. For the sites under consideration, the majority of the variability in the eddy-averaged winds is on the synoptic timescale, dominating influences from local features and large-scale circulation modes. Hence the large difference between daily and monthly (or seasonal) averaged quantities of the ratio A (Figure 4.10) and of θ (Figure 4.3). Conversely, the similarity of the monthly and seasonal averaged quantities of the ratio A and θ illustrate that there is little additional variability resolved at the monthly timescale versus at the seasonal timescale.

The fact that the mean vector wind components can be well predicted while μ is poorly predicted implies that high values of θ are not necessarily an a priori indication of good predictability of $\overline{w}$. However, the winds at the sites considered have predominantly small θ values; the 25th and 75th percentiles are 0.23 and 0.43 respectively. At sites with high θ winds, the relationship between μ and the mean vector wind could be robust, allowing for a robust μ predictability.

4.4 Returning to the four-variable Gaussian model

The above analysis has demonstrated that good predictability of $\overline{u}$ and $\overline{v}$ does not necessarily result in good predictability of μ. This leads to the conclusion that the IPM cannot be used as an a priori estimate of wind speed predictability at our sites. Nevertheless, if we return to considering all four of $\overline{u}$, $\overline{v}$, σ_u, and σ_v, it may still be the case that some information relevant to the predictability of $\overline{w}$ can be related to the sensitivity of $\overline{w}$ to vector wind statistics.

The dependence of the mean wind speed on the vector wind statistics may be quantified by calculating the direct correlation of $\overline{w}$ with means and standard deviations of the vector wind projections along various axes. While this approach is inspired by the idealized model, the calculated $\overline{w}$ sensitivities are model-independent and no assumptions need be made regarding the vector winds.

Figure 4.11 displays the calculated direct correlations (ρ) of Calgary's wintertime seasonal mean wind speed to the statistics of vector wind components (solid curves: means; dashed curves: standard deviations). The correlations are calculated for each of the 36 wind vector projections around the compass. From Figure 4.11, it is evident that while the fluctuations of the mean vector wind exhibit a moderate influence over seasonally-averaged $\overline{w}$ at Calgary, the mean wind speed variability is more strongly influenced by that of the standard deviation of the vector winds. A large majority of the variance ($\simeq 90\%$) in Calgary's DJF seasonal $\overline{w}$ may be explained by the fluctuations of the standard deviation of the south-easterly (north-westerly) vector wind. In contrast, the DJF seasonal $\overline{w}$ shares about half of its variance with the north-westerly (south-easterly) mean vector wind. Note that for this location the strongest correlation of $\overline{w}$ with the vector wind means and with the vector wind standard deviations occur along approximately orthogonal projections (of the vector wind), this is not always the case (as is evident in the plots for other locations in Appendix B).

The relatively high correlation of $\overline{w}$ to the vector wind standard deviations indicates that variability of the DJF seasonal $\overline{w}$ at Calgary is predominantly influenced by sub-seasonal variability. The weaker correlation of $\overline{w}$ to the vector wind means indicates that variability on seasonal timescales play a lesser role but nevertheless

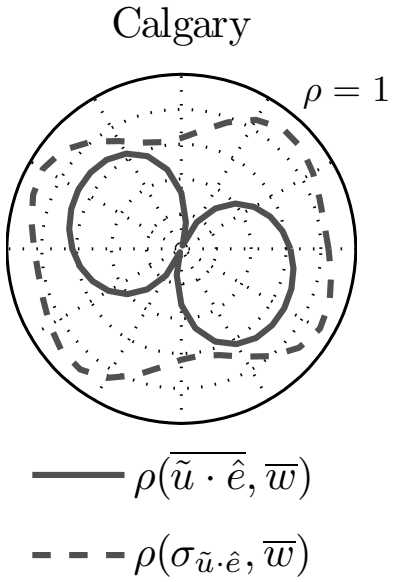

Figure 4.11: The correlation (ρ) of the DJF seasonal mean wind speed with the vector wind means (solid line), and with the vector wind standard deviations (dashed line). The outer circle denotes a reference ρ of unity. Each successive inner dotted circle corresponds to a change in ρ of 0.2.

imparts some influence on $\overline{w}$.

These polar correlation plots can be used to shed light on the polar prediction plots discussed in Chapter 3. In Figure 4.12, polar correlation plots of the winter-time monthly surface winds at six Albertan sites are displayed (top row) along with the corresponding polar prediction skill plots (bottom row). The sensitivity of the monthly mean wind speed to the statistics of the vector winds varies from site to site, although at most sites this dependence is strongly anisotropic. At nearly every site, season, and averaging timescale, there is a particular direction where the mean wind speed is insensitive to the variability of the mean vector wind. The anisotropy in the sensitivity of $\overline{w}$ variability to the vector wind standard deviations is not as strong.

A comparison of the polar correlation plots (Figure 4.12; top row) to the prediction plots (Figure 4.12; bottom row) demonstrates that mean wind speed predictability varies depending on the alignment between the vector wind component that is best predicted and that to which the mean wind speed is most sensitive. For example, the correlation plots for Lethbrige and Grande Prairie clearly show that the mean wind

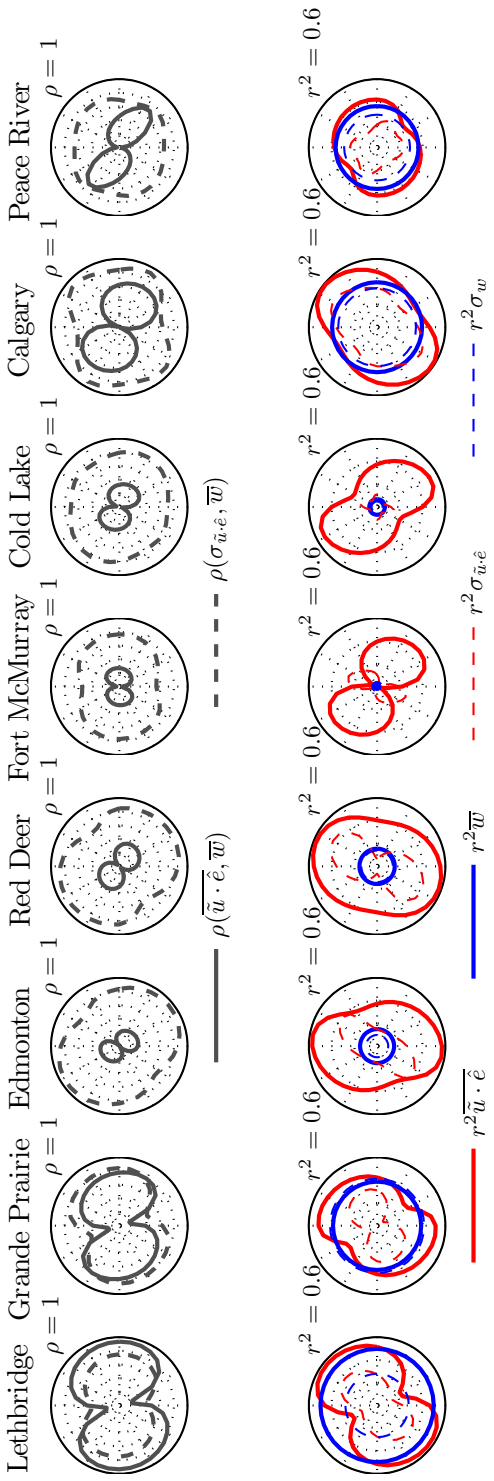

Figure 4.12: The correlation between the DJF monthly mean wind speed and the DJF monthly statistics of the vector projections (top row) at six Alberta sites, and the corresponding cross-validated r^2 prediction skills (bottom row).

speed is strongly correlated to the mean and standard deviation of the zonal winds. These strong sensitivities indicate that good predictability of statistical features of the zonal winds will result in good predictability of the mean wind speed. This is in fact evident in the corresponding polar prediction plots. At Lethbridge, there is good alignment between the directions of mean wind speed sensitivity and the predictability of vector wind statistics, with correspondingly good predictions of $\overline{w}$. At Grande Prairie, the alignment between the direction of the greatest wind speed sensitivity and the predictability of the mean vector winds is weaker than at Lethbridge. Consequently, the predictability of $\overline{w}$ at Grande Prairie is understandably reduced compared to that of Lethbridge.

Considering now the next four sites (Edmonton, Red Deer, Fort McMurray, and Cold Lake) in Figure 4.12, it is seen that there is very little shared variance between $\overline{w}$ and the mean of the vector projections, strongly suggesting that $\overline{w}$ is likely to be poorly predicted. This result is confirmed from the prediction skill results. The r^2 of $\overline{w}$ is somewhat higher at Edmonton and Red Deer ($r^2 \simeq 0.2$) than at Fort McMurray or Red Deer. In Edmonton and Red Deer, the sensitivities of $\overline{w}$ to $\sigma_{\widetilde{u}\cdot\hat{e}}$ are larger than those of Fort McMurray and Cold Lake. Note that since the vector wind standard deviations are generally poorly predicted, a strong sensitivity to this statistical feature alone does not not generally imply good wind speed predictability. However, at Edmonton and Red Deer, the standard deviations display relatively high degrees of predictability. Evidently, there is some information in the flow aloft relevant to the sub-monthly variability and therefore the monthly mean wind speed. The wind speed predictability is nevertheless small.

These polar plots represent a more general means of interpreting mean wind speed predictability in terms of the statistical features of the vector winds than that provided by the IPM. In particular, they demonstrate that good mean wind speed predictions do not result from good predictions of the vector wind components alone; alignment between directions of good predictability and directions of $\overline{w}$ sensitivity is also important.

Chapter 5

Conclusions

This study's purpose was two-fold. First, the study sought to quantify the predictability of vector and scalar surface winds at Canadian sites with simple statistical downscaling methods. Second, the study sought to investigate the relative predictability of scalar and vector wind quantities.

5.1 Summary of Results

The predictive information contained in the flow aloft relevant to the surface winds at sites in Alberta, Saskatchewan, and Ontario was found to occur on large scales throughout the mid-troposphere. The variability aloft was decomposed into Empirical Orthogonal Functions. The leading principal components, which carry the large-scale information aloft and thus much if not all of the predictive information, were used to drive a statistical downscaling model. The linear multivariate regression model made predictions of statistical features of the historical wind observations on daily, monthly, and seasonal averaging timescales. Predictions were made for each of the three-month calendar seasons; DJF, MAM, JJA, and SON.

In general, the statistical features of the wind speed are poorly predicted. More skillful predictions were obtained of the mean vector winds, and to a lesser degree the vector wind standard deviations. Vector wind predictability was found to be

highly anisotropic. Typically on all averaging timescales there is a mean vector wind component with prediction r^2 in excess of 0.5.

A two-variable idealized probability model (IPM) of the mean wind speed was used to understand the differences in predictability between wind speed statistics and vector wind statistics. The IPM assumed that vector wind fluctuations are isotropic, uncorrelated, and Gaussian. Empirical evidence was provided that this model is able to provide an excellent characterisation of the relationship between mean wind speed and statistics of the vector winds. Analysis of the IPM demonstrated that mean wind speed predictability can be explained by the predictability of the mean vector wind magnitude and of the vector wind's standard deviation. In particular, the analysis demonstrated two end-member regimes of wind speed sensitivities. In one regime (low θ), the predictability of the mean wind speed may be explained solely by the predictability of the standard deviation of the vector winds (Figure 5.1; $r^2\,\sigma$). In the other (high θ), the mean wind speed predictability is dependent on the predictability of the mean vector wind magnitude (Figure 5.1; $r^2\,\mu$)).

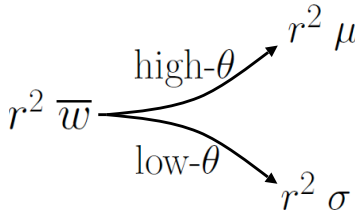

Figure 5.1: A schematic relating mean wind speed predictability to that of the mean vector wind magnitude (μ; high θ) and that of the vector wind's standard deviation (σ; low θ). Between these is a continuum of moderate θ for which mean wind speed predictability will not depend solely on one of μ and σ.

The utility of these mean wind speed sensitivities as an a priori estimate of mean wind speed predictability was explored. For the winds under consideration, neither the magnitude of the mean vector winds nor the vector wind's standard deviation was found to possess an inherent predictability similar to that of the mean vector wind components. The predictability of the vector wind's standard deviation is generally poor, while the potential for high predictability of the mean vector wind magnitude was shown to depend on the strength of the prevailing winds. When the prevailing winds dominate (i.e. Figure 5.2, $mean(\overline{\tilde{u}\cdot\hat{e}}) \gtrsim std(\overline{\tilde{u}\cdot\hat{e}})$), the potential predictability of the mean vector wind magnitude ($r^2\,\mu$) will approach that of the mean vector wind

components. When the prevailing winds are weak (i.e. Figure 5.2, $mean(\overline{\tilde{u}\cdot\hat{e}}) < std(\overline{\tilde{u}\cdot\hat{e}})$), the ceiling of $r^2\,\mu$ is considerably lower than that of the mean vector wind components. We thus find high θ to be a necessary, but not sufficient condition for high mean wind speed predictability.

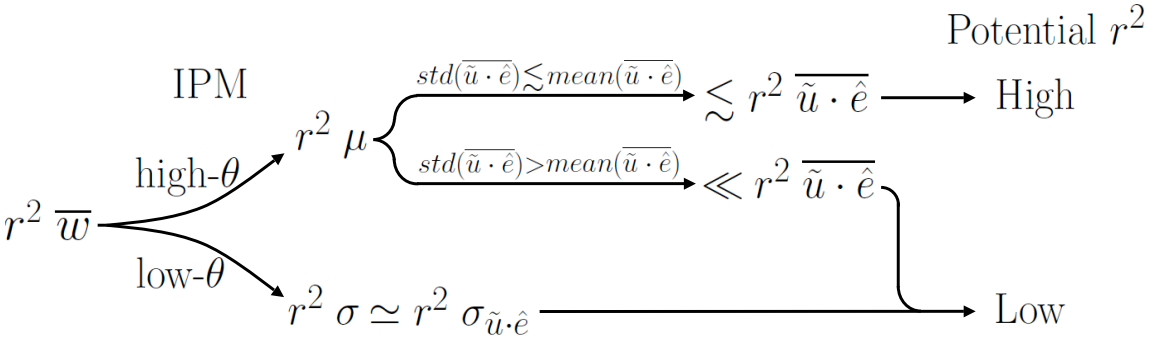

Figure 5.2: A schematic depicting the controls on mean wind speed predictability relative to the predictability of the vector winds. Mean wind speed predictability is explained by that of μ in high θ conditions, and that of σ in low θ conditions. The predictability of σ is approximately equal to that of the vector wind's standard deviations, which is generally low. The predictability of μ is generally less than that of the mean vector wind components, however, if the prevailing winds dominate (i.e. $std(\overline{\tilde{u}\cdot\hat{e}}) < mean(\overline{\tilde{u}\cdot\hat{e}})$), μ's predictability will approach that of the mean vector winds. On the other hand, if the prevailing winds are weak relative to their variability, μ's predictability will generally be much lower than that of the mean vector winds . Consequently, mean wind speed has a potentially high predictability only with a) high θ conditions, and b) strong prevailing winds; otherwise, mean wind speed predictability is expected to be low.

A heuristic relation between mean wind speed predictability and that of the vector wind quantities was also found. Polar correlation plots of wind speed sensitivities to statistics of the vector wind, at projections around the compass, demonstrated that good predictions of the mean wind speed require that vector wind fluctuations in the direction to which the mean wind speed is most sensitive must themselves be well predicted.

5.1.1 Systematic Features of Statistical Predictability of Surface Winds

The results of this study demonstrate a high degree of variability in the predictability of wind statistics across different sites and seasons. Some general patterns do emerge from this variability:

1. the means are generally better predicted than the standard deviations, for both scalar and vector predictands;
2. the vector predictands are generally better predicted than the scalar predictands;
3. there is in general a strong anisotropy in the predictabilty of the vector winds;
4. predictive information is common to mean wind speed and the standard deviation of wind speed;
5. predictive information relevant to mean vector winds is distinct from that of the vector wind standard deviations; and lastly,
6. the prediction skill of mean quantities generally increases with decreases in the averaging timescale (from seasonal to monthly to daily).

5.2 Discussion of Results

The predictability of the surface winds at two southern Alberta and four southern Saskatchewan sites was explored, to a limited extent, in St. George and Wolfe (2009). The findings of this earlier study motivated the selection of some of this study's sites. St. George and Wolfe (2009) found a direct correlation (ρ) of 0.49 between regional wintertime seasonal wind speeds in southern Canadian prairies (SCP) and the Niño 3.4 index. However, as St. George and Wolfe did not perform a cross-validated prediction, this correlation likely represents an upper bound on the linear predictive skill of the Niño 3.4 index. Furthermore, the regional wind speed time series in St. George and Wolfe (2009) was constructed by averaging the wintertime seasonal means of six widely-separated sites. In the present study, predictions were cross-validated and sites were predicted individually (thus including local effects that may

have been lost through aggregation). Nevertheless, the predictability of the SCP mean winds was greater than that indicated by direct correlations with the Niño 3.4 index. Across the SCP, the mean wintertime wind speed prediction r^2 was 0.31, in contrast to 0.24 from Niño 3.4. Despite this improvement over the estimate of prediction skill reported by St. George and Wolfe, these cross-validated speed prediction skills are still relatively small.

One potential source of error in our prediction results is the construction of the predictands from imperfect data. The direct correlations (ρ; across site and season) of the homogenized mean wind speeds (Wan et al., 2010) and means computed directly from the adjusted hourly data provide a measure of the impact of the discontinuities that remain in the adjusted data. The histogram in Figure 5.3 illustrates the distribution of correlations of the seasonally-averaged means; they ranged from 0.39 to 1.0, with the lowest correlations typically found in the summer calendar season. The agreement between the monthly means was noticeably improved; the lower bound of the correlations increased to 0.6 (not shown). The varying import of the discontinuities represents an added difficulty in directly comparing predictions on daily averaging timescales with those on longer timescales.

The scatterplot in Figure 5.3 displays the relationship between seasonal mean wind speed predictability (by site and season) and the corresponding correlation of the homogenized means and those computed directly from the adjusted hourly data. There is clearly no linear relationship between prediction skill and the agreement of the means from the two datasets. Since predictability is expected to be on average negatively affected by these non-stationarities, the study's results represent a "lower bound" on SD's achievable results.

The systematic results we found across sites can be understood in terms of the IPM and the results of previous studies.

1. The predictability of the means versus standard deviations

The standard deviations of wind quantities are generally not well predicted, while the means are. The averaging operation partitions variability into resolved components (the mean) and unresolved components (the sub-averaging timescale standard

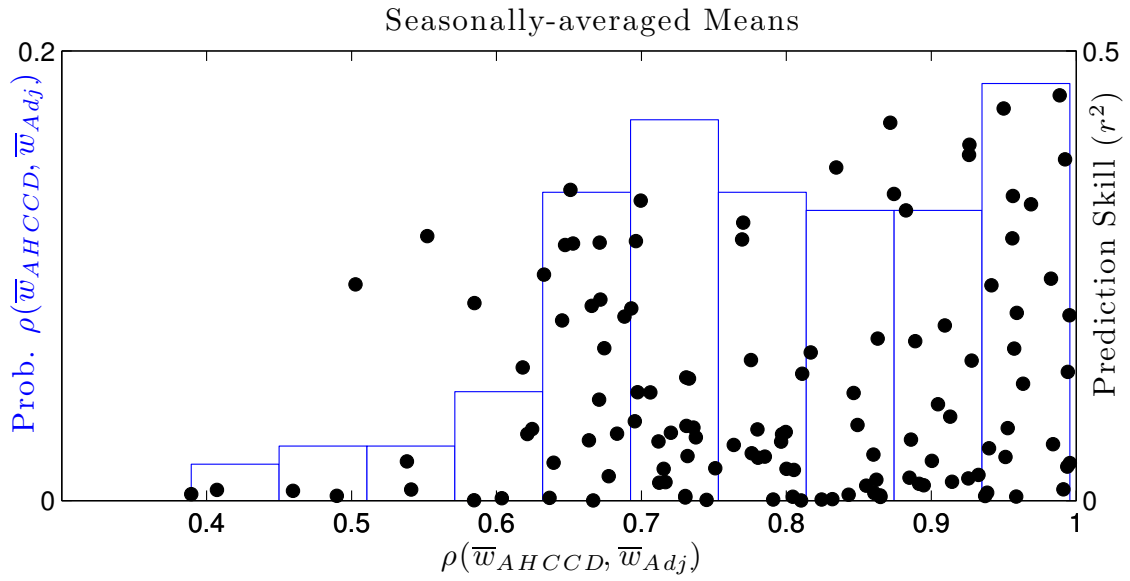

Figure 5.3: The frequency distribution of the direct correlation (across site and season) of seasonal mean wind speed quantities from the homogenized dataset and quantities computed from the adjusted dataset (blue histogram). A scatterplot of mean wind speed predictability relative to the direct correlation of the homogenized and adjusted mean wind speed quantities overlays the frequency distribution.

deviations). This systematic result indicates that the statistical downscaling model best predicts the resolved variability, and that the large-scale averaged flow does not carry useful predictive information for the sub-averaging timescale standard deviations. This is also reflected in the mean wind speed predictability: the mean speed of low θ winds is generally poorly predicted, while the mean wind speed at high θ sites was typically associated with the highest prediction skill.

2. The superior predictability of vector wind means relative to the scalar winds

There is not a large body of work systematically comparing the predictability of vector winds to that of wind speed. In the few studies that have been carried out, relatively high vector wind predictability is consistently found (e.g. Salameh et al., 2009; van der Kamp et al., 2011; Monahan, 2012a). The IPM demonstrates that when the vector wind component's standard deviations are relatively large (which is typical of longer averaging timescales), their predictability will generally determine

that of the mean wind speed - regardless of how well the mean vector winds are predicted. Given that the standard deviations of the vector winds are not generally well-predicted, it follows that the scalar quantities will not be well predicted. On the daily averaging timescale, for which the vector wind standard deviations are relatively small, the IPM demonstrates that mean wind speed predictability generally depends equally on the magnitude of the vector wind means. At this averaging timescale, when the fluctuations of the resolved variability (the means) dominates fluctuations of the unresolved variability, mean wind speed predictability is relatively high.

We also note that with vector wind components, there is generally a vector projection that maximizes the synchronization of fluctuations in the predictors and in the predictands. This anisotropy is not a feature of the scalar wind quantities.

3. The anisotropy of prediction skill

Two earlier studies, van der Kamp et al. (2011) and Salameh et al. (2009), have directly addressed the anisotropy in the vector winds' predictabilty. Both attribute the anisotropy to the influence of topography because of the observed generally strong alignment of predictability along topographic features. The van der Kamp et al. (2011) study found alignment of predictability with strong local and, on occasion, more remote topographic features at most sites. There were also a small number of sites whose vector wind predictability did not align with local or mesoscale features. It is clear from Figure 3.4 of the current study that sites of similar distances from the Rockies exhibit contrasting orientations of anisotropy in mean vector wind predictability. Furthermore, sites in regions that are devoid of large topographical features, such as the southern prairies, exhibited equally strong anisotropy as sites with strong local topography. These results suggest that while topography may be an important factor, it is not the sole influence on the anisotropy of predictability.

Salameh et al. (2009) suggest that along-topography winds are better predicted because their variability is on lower frequencies (and therefore with longer autocorrelation times) than across-topography winds. Topography is not the only influence that can produce anisotropy in vector wind fluctuation timescales. While at many locations the dominant constraint will likely be the surrounding topography, at other locales, such as the southern prairies, the main influence may be the dominant tracks

of cyclonic and anti-cyclonic activity and fronts. Monahan (2012c) has demonstrated that anisotropy in the kinematics of vector wind fluctuations in propagating synoptic eddies results in anisotropy of the one-day lag vector wind autocorrelation. The surface vector wind lag autocorrelation structure could be a convenient tool to understand the origin of the anisotropy of vector wind predictability.

4. and 5. The shared predictive information across the statistical features of the surface winds

Assuming that both of the IPM's two-variables, μ and σ, are of comparable variance, the sensitivities $\partial_\mu \overline{w}$, $\partial_\sigma \overline{w}$, $\partial_\mu \sigma_w$, and $\partial_\sigma \sigma_w$, demonstrate that systematic result 4 only holds for low θ conditions for which the variability of the mean and the standard deviation of the wind speed are dominated by the same model parameter: the isotropic standard deviation of the vector winds.

Systematic result 5 indicates that little predictive information is held in common between the mean vector winds and the standard deviation of the vector winds. As such, other predictors and likely other models will be necessary to successfully predict the standard deviations. An exploratory correlation analysis demonstrated that the standard deviation of the predictor fields aloft are not suitable predictor candidates for the standard deviations of surface vector winds; an important direction of future study is the investigation of other ways to predict these standard deviations.

6. The increasing predictability with decreasing averaging timescales

Two earlier studies have investigated the dependence of prediction skill to the averaging timescales: Salameh et al. (2009) and Monahan (2012a). Salameh et al. (2009) made predictions of 6-hourly, daily, and weekly mean surface flows at six French cities. At these sites, prediction skill of vector components was found to increase with the averaging timescale. On the other hand, Monahan (2012a) (and the current study) have found predictability to be best with the shortest averaging timescales. There are limitations to comparisons made between the prediction skill found in Salameh et al. (2009), Monahan (2012a), and the current study. First, the daily averaging timescale is the only averaging timescale held in common by all three of these studies. Sec-

ond, Salameh et al. (2009) focused on highly topographically-forced sites; Monahan focused on sea surface winds in the Sub-arctic Pacific, in the absence of topographic forcing. Local topographic influences on winds in Alberta, Saskatchewan, and Ontario would fall somewhere between the end member cases studied by Salameh and Monahan. Third, in the earlier two studies, only the zonal and meridional components were predicted. The van der Kamp et al. (2011) study found the direction of the best predicted component varied by up to 30 degrees from one calendar season to another and was not in general aligned zonally or meridionally. In the current study, the direction of the best predicted component was also found to vary with the averaging timescale. The anisotropy of the vector wind predictability in Salameh et al. (2009) and Monahan (2012a) is an added uncertainty in their observed response of predictability to the averaging timescale. These uncertainties suggest caution in directly comparing the observed influence of the averaging timescale on predictability across these studies.

Nevertheless, these studies suggest a non-linear response in predictability to a change in averaging timescale. This is expected as the variability resolved by the means of surface winds varies non-linearly with the averaging timescale. We have seen that resolved variability in vector winds is well predicted by these predictors and models, but unresolved variability is not. On sub-daily timescales, variability is largely locally controlled and therefore not well predicted by large-scale flow. On monthly and seasonal timescales, the dominant synoptic variability has been averaged out and again the predictions are poor. Prediction skill is particularly poor on seasonal timescales, as the absolute variability of the means is low. On daily timescales, local effects have been suppressed but the dominant synoptic day-to-day variability remains. These results suggest that when the dominant variability is resolved (by the predictor(s) and predictand(s)), prediction skill is at its best. Developing a deeper understanding of the controls on the relative predictability of winds across different averaging timescales is an important direction for future research.

5.3 Broader Implications of Results

The statistical methods considered in the present study are found to yield only weak predictability of quantities heavily influenced by sub-averaing timescale variability;

the monthly- and longer mean land surface wind speeds, as well as the standard deviations of the scalar and vector winds. The relationship between the state of the troposphere averaged over a particular timescale and the sub-averaging timescale variability of the surface winds is either more complex than the capabilities of linear statistical models or non-existent. If a more complex relationship does exist, successfully modeling it with statistical downscaling methods requires additional model parameters and possibly non-linear models. Unfortunately, with the statistical degrees of freedom in current observational records, incorporating additional model parameters may come at the expense of model robustness. This result suggests that deterministic models, which resolve the short-timescale variability, may be the most useful tool for successful predictions of wind quantities that are highly influenced by sub-averaging timescale variability.

The prediction results were much improved when predicting wind quantities on timescales such that the dominant variability is resolved, particularly so with mean wind speed quantities at times and places where the prevailing winds dominate. Further evaluations of mean wind speed predictability at additional high θ sites would be needed to assess the generality of this finding. As land surface winds are influenced by many factors, such an investigation of wind predictability in high θ regions would likely be best served by a focus on sea surface trade winds, where strong mean winds dominate with weak variability. Nevertheless within British Columbia, there are wind farm sites in the Peace river region where anecdotal evidence suggests that the surface winds are dominated by the prevailing winds (Kroeker, 2011). Thus, these winds are potentially in a high θ regime, and are also ideal candidates for high wind speed predictability.

The present study has demonstrated a theoretical and empirical relation of mean wind speed predictability and the predictability of the vector winds. Seemingly random quantitive differences between the two have been shown to be a result of the varying influence of the vector wind quantities on the scalar winds. The results suggest that predictions of scalar wind standard deviations, as well as mean winds in low θ conditions, made with statistical downscaling models are unlikely to be successful. For applications that require fore-knowledge of variability in the wind resource, such as wind power forecasting, civil engineering and design considerations, as well as transport and dispersion models, etc., these results suggest the potential importance

of dynamic downscaling for surface winds. Furthermore, these results emphasize that information regarding the mean vector winds may have little relevance to the mean wind speeds; particularly in low θ conditions, as the mean vector winds will have no bearing on wind speed quantities.

Bibliography

Allan, R., J. Lindesay, and D. Parker, 1996: *El Nino Southern Oscillation and Climatic Variability*. CSIRO Publishing, first edition.

BC Hydro, 2011: Clean Power Call.
URL http://www.bchydro.com/planning_regulatory/acquiring_power/clean_power_call.html?WT.mc_id=rd_cleanpowercall

Braxmeier, H., 2011: .
URL http://www.maps-for-free.com/index/

Buell, E., 1979: On the physical interpretation of empirical orthogonal functions. *Preprints, Sixth Conference on Probability and Statistics in Atmospheric Sciences*, American Meteorological Society, Banff AB, Canada, 112–117.

Bukovsky, M. S. and D. J. Karoly, 2007: A brief evaluation of precipitation from the North American Regional Reanalysis. *Journal of Hydrometeorology*, **8**, 837–846.
URL http://dx.doi.org/10.1175/JHM595.1

Canada, E., 1977: *Manual of surface weather observations*. 7th ed. 419 pp. edition.

— 2011: National climate data and information archive: Climate data online.
URL http://climate.weatheroffice.gc.ca/climateData/canada_e.html

Curry, C., D. van der Kamp, and A. Monahan, 2011: Statistical downscaling of historical monthly mean winds over a coastal region of complex terrain. I. Predicting wind speed. *Climate Dynamics*, 1–19, 10.1007/s00382-011-1173-3.
URL http://dx.doi.org/10.1007/s00382-011-1173-3

Davy, R., M. Woods, C. Russell, and P. Coppin, 2010: Statistical downscaling of wind variability from meteorological fields. *Boundary-Layer Meteorology*, **135**, 161–175,

10.1007/s10546-009-9462-7.
URL http://dx.doi.org/10.1007/s10546-009-9462-7

Giorgi, F. and L. O. Mearns, 1991: Approaches to the simulation of regional climate change: A review. *Rev. Geophys.*, **29**, 191–216.
URL http://dx.doi.org/10.1029/90RG02636

Goubanova, K., V. Echevin, B. Dewitte, F. Codron, K. Takahashi, P. Terray, and M. Vrac, 2011: Statistical downscaling of sea-surface wind over the Peru–Chile upwelling region: diagnosing the impact of climate change from the IPSL-CM4 model. *Climate Dynamics*, **36**, 1365–1378, 10.1007/s00382-010-0824-0.
URL http://dx.doi.org/10.1007/s00382-010-0824-0

He, Y., A. H. Monahan, C. Jones, A. Dai, D. Caya, S. Biner, and K. Winger, 2010: Land surface wind speed probability distributions in North America: Observations and regional climate model evaluations. *Journal of geophysical research*, **115**.

Holton, J., 2004: *An Introduction to Dynamic Meteorology*, volume 88 of *International Geophysics Series*. Academic Press, fourth edition.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph, 1996: The NCEP/NCAR 40-year reanalysis projectear Reanalysis Project. *Bulletin of the American Meteorological Society*, **77**, 437–471.
URL http://dx.doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Kanamaru, H. and M. Kanamitsu, 2007: Scale-selective bias correction in a downscaling of global analysis using a regional model. *Monthly Weather Review*, **135**, 334–350.
URL http://dx.doi.org/10.1175/MWR3294.1

Kroeker, M., 2011: personal communication.

Mesinger, F., G. Dimego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jovi, J. Woollen, E. Rogers, E. H. Berbery, M. B. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Lin, G. Manikin, D. Parrish, and W. Shi, 2006: North

American Regional Reanalysis. *Bulletin of the American Meteorological Society*, **87**, 343–360, doi:10.1175/BAMS-87-3-343.

Monahan, A. H., 2012a: Can we see the wind? Statistical downscaling of historical sea surface winds in the subarctic northeast Pacific. *Journal of Climate*, **25**, 1511–1528.
URL http://dx.doi.org/10.1175/2011JCLI4089.1

— 2012b: Limits to the gaussian predictability of wind speeds, submitted to Journal of Climate.

— 2012c: The temporal autocorrelation structure of sea surface the temporal autocorrelation structure of sea surface winds, submitted to Journal of Climate.

Najac, J., J. Boé, and L. Terray, 2009: A multi-model ensemble approach for assessment of climate change impact on surface winds in France. *Climate Dynamics*, **32**, 615–634, 10.1007/s00382-008-0440-4.
URL http://dx.doi.org/10.1007/s00382-008-0440-4

Ontario Power Authority, 2010: FIT program.
URL http://fit.powerauthority.on.ca/

Pryor, S. C., J. T. Schoof, and R. J. Barthelmie, 2005: Empirical downscaling of wind speed probability distributions. *J. Geophys. Res.*, **110**.
URL http://dx.doi.org/10.1029/2005JD005899

Robeson, S. and K. Shein, 1997: Spatial coherence and decay of wind speed and power in the north-central united states rid a-9895-2008. *Physical Geography*, **18**, 479–495.

Salameh, T., P. Drobinski, M. Vrac, and P. Naveau, 2009: Statistical downscaling of near-surface wind over complex terrain in southern france. *Meteorology and Atmospheric Physics*, **103**, 253–265, 10.1007/s00703-008-0330-7.
URL http://dx.doi.org/10.1007/s00703-008-0330-7

St. George, S. and S. A. Wolfe, 2009: El Nino stills winter winds across the southern Canadian prairies. *Geophysical Research Letters*, **36**.

Uppala, S., P. Kallberg, A. Simmons, U. Andrae, and V. Bechtold, 2005: The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, **131**, 2961–3012.

van der Kamp, D., C. Curry, and A. Monahan, 2011: Statistical downscaling of historical monthly mean winds over a coastal region of complex terrain. II. Predicting wind components. *Climate Dynamics*, 1–11.
URL http://dx.doi.org/10.1007/s00382-011-1175-1

Wan, H., X. L. Wang, and V. R. Swail, 2010: Homogenization and Trend Analysis of Canadian Near-Surface Wind Speeds. *Journal of Climate*, **23**, 1209–1225.
URL http://dx.doi.org/10.1175/2009JCLI3200.1

Wang, X. L., 2008: Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or f test. *Journal of Applied Meteorology and Climatology*, **47**, 2423–2444.
URL http://dx.doi.org/10.1175/2008JAMC1741.1

White, G., M. Fiorino, M. Kanamitsu, S. Saha, E. Kalnay, V. Kousky, R. Kistler, W. Collins, R. Jenne, J. Woollen, M. Chelliah, and W. Ebisuzaki, 2001: The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society*, **82**, 247–267.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, San Diego, 2 edition.

Zorita, E. and H. von Storch, 1999: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *Journal of Climate*, **12**, 2474–2489.
URL http://dx.doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2

Appendix A

Glossary

Table A.1: Explanations of frequently used terms.

Term	Explanation
scalar wind	the speed of the surface wind (no information of wind's direction)
vector wind	any vector projection of the surface wind (i.e. the zonal wind is a westerly or easterly projection of the surface wind)
surface wind quantity	either a mean or standard deviation of the surface wind speed (i.e. scalar wind quantity) or vector component (i.e. vector wind quantity)
daily-, monthly-, seasonal-mean	the mean of hourly observations computed within a day, month, or season
daily-, monthly-, seasonal-standard deviation	the standard deviation of hourly observations within a day, month, or season
vector wind means	synonomous with mean vector wind and refers to the means of a vector projection of the surface wind
statistical features of the surface wind	the mean and standard deviation (on a particular averaging timescale) of the scalar or vector wind
predictor variable	a 4-D meteorological field from the reanalysis products
predictor	a principal component of the extended-EOF of the five predictor variables
predictability	the general cross-validated prediction skill of a surface wind quantity
magnitude of the mean vector winds	the amplitude of two orthogonal mean vector winds, $\sqrt{\overline{u}^2 + \overline{v}^2}$

Table A.2: Explanations of frequently used symbols.

Symbol	Explanation
$\overline{u}$, $\overline{v}$	used to denote two orthogonal mean vector wind components
$\overline{\tilde{u} \cdot \hat{e}}$	the mean of the surface wind vector component in the direction $\hat{e}$
$\sigma_{\tilde{u} \cdot \hat{e}}$	the standard deviation of the surface wind vector component in the direction $\hat{e}$
$\overline{w}$	the mean of the surface wind speed
σ_w	the standard deviation of the surface wind speed
μ	the amplitude of two orthogonal mean vector winds, $\sqrt{\overline{u}^2 + \overline{v}^2}$
σ	an isotropic estimate of the vector wind standard deviation
θ	a ratio of the variability occuring on timescales equal to or greater than the averaging timescale relative to that on the sub-averaging timescale, defined as $atan(\mu/\sigma)$
mean$(\overline{\tilde{u} \cdot \hat{e}})$	the long-term mean of the (daily, monthly, or seasonally-varying) mean vector wind
std$(\overline{\tilde{u} \cdot \hat{e}})$	the long-term standard deviation of the (daily, monthly, or seasonally-varying) mean vector wind

Appendix B

Station Information

Table B.1: The adjusted surface wind stations used in this study; all are located at airports.

Province	Station Name	Station ID	Beg. Yr/Mth	End Yr/Mth	Lat. (deg)	Long. (deg)	Elev. (m)
AB	Calgary Int'l	3031093	1953/1	2006/12	51.11	-114.02	1084
AB	Cold Lake	3081680	1954/8	2006/12	54.42	-110.28	541
AB	Edmonton City	3012208	1953/1	2004/12	53.57	-113.52	671
AB	Edmonton Int'l	3012205	1961/1	2006/12	53.32	-113.58	723
AB	Fort McMuray	3062693	1953/1	2006/12	56.65	-111.22	369
AB	Grande Prairie	3072920	1953/1	2006/12	55.18	-118.89	669
AB	Lethbridge	3033880	1953/1	2006//12	49.63	-112.8	929
AB	Medicine Hat	3034480	1953/1	2006/12	50.02	-110.72	717
AB	Peace River	3075040	1955/1	2006/12	56.23	-117.45	571
AB	Red Deer	3025480	1953/1	2006/12	52.18	-113.89	905
SK	Estevan	4012400	1953/1	2006/12	49.22	-102.97	581
SK	Moose Jaw	4015320	1954/1	2006/12	50.33	-105.55	577
SK	Regina Int'l	4016560	1953/1	2006/12	50.43	-104.67	577
SK	Swift Current	4028040	1953/1	2006/12	50.3	-107.68	818
ON	Earlton	6072225	1953/1	2004/12	47.7	-79.85	243
ON	Gore Bay	6092925	1953/1	2005/12	45.88	-82.57	194
ON	Kapuskasing	6073975	1953/1	2006/12	49.41	-82.47	227
ON	Kenora	6034075	1953/1	2006/12	49.79	-94.37	410
ON	London Int'l	6144475	1953/1	2006/12	43.03	-81.15	278

Continued on next page

Table B.1 – continued from previous page

Province	Station Name	Station ID	Beg. Yr/Mth	End Yr/Mth	Lat. (deg)	Long. (deg)	Elev. (m)
ON	Muskoka	6115525	1953/1	2004/12	44.97	-79.3	282
ON	North Bay	6085700	1953/1	2006/12	46.36	-79.42	370
ON	Ottawa M-C Int'l	6106000	1953/1	2006/12	45.32	-75.67	114
ON	Sault Ste Marie	6057592	1961/8	2006/12	46.48	-84.51	192
ON	Sudbury	6068150	1954/1	2006/12	46.63	-80.8	348
ON	Thunder Bay	6048261	1953/1	2006/12	48.37	-89.33	199
ON	Timmins V. Power	6078285	1955/4	2006/12	48.57	-81.38	295
ON	Toronto Island	6158665	1957/2	2006/12	43.63	-79.4	77
ON	Toronto YYZ	6158733	1953/1	2006/12	43.68	-79.63	173
ON	Trenton	6158875	1953/1	2006/12	44.12	-77.53	86
ON	Wiarton	6119500	1953/1	2006/12	44.75	-81.11	222
ON	Windsor	6139525	1953/1	2006/12	42.28	-82.96	190

Table B.2: Station Metadata

Station ID	Start (yyyymmdd)	End (yyyymmdd)	Hgt. (m)	Anemometer Type
3012205	19590624	19630222	0	DYNES
3012205	19630222	19700199	10	45B
3012205	19630222	99999999	10	U2A/R
3012208	19550100	19690700	18.2	45B
3012208	19650727	19690700	18.2	U2A
3012208	19690700	19700120	17.3	U2A
3012208	19700120	99999999	10	U2A/R
3025480	19391200	19610700	15.5	45B
3025480	19610700	19710409	18.2	45B
3025480	19610800	19710400	18.2	U2A
3025480	19710400	19810331	17.7	U2A
3025480	19810331	99999999	10	U2A
3031093	19231123	19480324	18	None
3031093	19480324	19530400	19.8	None
3031093	19530400	19560200	16.7	DYNES
3031093	19560200	19680417	20.4	DYNES
3031093	19560200	19680417	20.4	45B
3031093	19560200	19730504	20.4	U2A
3031093	19730504	19861008	18.3	U2A/R
3031093	19861009	99999999	10	U2A/R
3033880	19480300	19531200	25.8	45B
3033880	19531200	19590800	19.1	45B
3033880	19590800	19651100	18.7	45B
3033880	19651100	19721299	10	45B
3033880	19651100	19730501	10	U2A
3033880	19730501	99999999	10	U2A/R
3034480	19401000	19430500	12.2	None
3034480	19430500	19450627	13.4	45B
3034480	19450627	19590806	17.4	45B
3034480	19590806	19651000	17.4	45B
3034480	19651010	19700727	15.2	U2A

Continued on next page

Table B.2 – continued from previous page

Station ID	Start (yyyymmdd)	End (yyyymmdd)	Hgt. (m)	Anemometer Type
3034480	19651000	19700799	0	45B
3034480	19700727	99999999	10	U2A
3062693	19470000	19491000	6.4	None
3062693	19491000	19520900	12.8	None
3062693	19520900	19630900	19.2	U2A
3062693	19630900	19760914	10	U2A
3062693	19760914	19760914	13.1	U2A
3062693	19760914	99999999	13.1	U2A/DIAL
3072920	19430300	19551100	18.3	45B
3072920	19551100	19631100	17.7	45B
3072920	19631100	19710199	10	45B
3072920	19631100	19631100	10	U2A
3072920	19631100	99999999	10	U2A/DIAL
3075040	19550500	19630222	11	None
3075040	19630222	99999999	10	U2A
3081680	19580200	19620900	21.9	U2A
3081680	19620900	19640600	21.3	U2A
3081680	19640600	19700700	10	U2A
3081680	19700700	99999999	10	U2A/R
4012400	19471100	19490500	19.8	45B
4012400	19490500	19621003	17.4	45B
4012400	19621003	19681219	10	45B
4012400	19681219	19770916	10	U2A
4012400	19770916	99999999	10	U2A/R
4015320	19471000	19510900	19.8	None
4015320	19510900	19560400	14.3	45B
4015320	19560400	19600100	31.7	45B
4015320	19600100	19640302	14.3	U2A
4015320	19640302	19710908	10	U2A
4015320	19710908	99999999	10	U2A
4016560	19410800	19430500	20.4	45B

Continued on next page

Table B.2 – continued from previous page

Station ID	Start (yyyymmdd)	End (yyyymmdd)	Hgt. (m)	Anemometer Type
4016560	19430500	19490500	21.3	45B
4016560	19560400	19601012	20.4	45B
4016560	19601013	19850221	26.8	45B
4016560	19670100	19680200	9.1	U2A
4016560	19680200	99999999	10	U2A/R
4028040	19380600	19390500	0	None
4028040	19390500	19521200	18.3	None
4028040	19521200	19660500	17	45B
4028040	19660500	19730927	13.1	45B
4028040	19670100	19740124	10	U2A
4028040	19740124	99999999	10	U2A/R
6034075	19380808	19430519	0	None
6034075	19430519	19510907	28.9	None
6034075	19510907	19560817	27.4	45B
6034075	19560817	19680402	26.1	45B
6034075	19680408	19700119	10	U2A
6034075	19710400	19841217	10	U2A/R
6034075	19841217	99999999	10	U2A
6048261	19410800	19430602	25.9	45B
6048261	19430602	19560118	24.3	45B
6048261	19560118	19620729	13.7	45B
6048261	19620729	19651031	11.3	45B
6048261	19651031	19830516	10	45B
6048261	19701022	99999999	10	U2A/R
6057592	19460417	19610728	6.1	45B
6057592	19610728	19680427	13.4	U2A
6057592	19680427	19701127	12.2	U2A
6057592	19701127	19731101	12.8	U2A
6057592	19731101	19781208	11.6	U2A
6057592	19781208	19811211	11.6	U2A/R
6057592	19811211	99999999	10	U2A/R
6068150	19540210	19580604	10.3	45B

Continued on next page

Table B.2 – continued from previous page

Station ID	Start (yyyymmdd)	End (yyyymmdd)	Hgt. (m)	Anemometer Type
6068150	19580604	19620813	13.7	45B
6068150	19620813	19701117	14.6	45B
6068150	19620813	19670602	14.5	U2A
6068150	19670602	19701117	10	U2A
6068150	19701117	19721110	14.7	45B
6068150	19701117	19721110	10.4	U2A
6068150	19721110	19820122	14.6	45B
6068150	19721110	99999999	10	U2A
6072225	19380908	19511200	19.8	45B
6072225	19511200	19630124	15.2	45B
6072225	19630124	19680506	19.2	U2A
6072225	19680506	99999999	19.3	U2A
6072225	19701117	99999999	19.5	U2A
6073975	19381207	19411107	12.8	45B
6073975	19590517	19620920	13.7	45B
6073975	19620920	19680424	13.7	45B
6073975	19620920	19680424	18.9	U2A
6073975	19680424	19691125	14.6	45B
6073975	19711026	19730000	18.8	U2A
6073975	19730000	19780830	19.3	U2A
6073975	19780830	19830331	9.7	U2A
6073975	19830331	99999999	10	U2A
6073975	19411107	19590517	12.2	45B
6078285	19550612	19621002	10.7	45B
6078285	19621002	19680430	12.3	U2A
6078285	19680430	19730119	10.8	U2A
6078285	19730119	19780106	9.9	U2A
6078285	19780106	19790224	10	U2A
6078285	19790224	99999999	9.9	U2A
6085700	19390100	19500719	15.2	45B
6085700	19500719	19631212	19.5	45B
6085700	19631212	19670602	12.6	BEND/FRIEZ

Continued on next page

Table B.2 – continued from previous page

Station ID	Start (yyyymmdd)	End (yyyymmdd)	Hgt. (m)	Anemometer Type
6085700	19670602	19691006	12.8	BEND/FRIEZ
6085700	19691006	19791130	13.1	U2A/R
6085700	19791130	99999999	13	U2A/R
6092925	19500000	19631030	13.1	45B
6092925	19631030	19830613	9.7	U2A
6092925	19830613	19830613	10	U2A
6092925	19830613	99999999	10	U2A/DIAL
6106000	19440519	19510210	18.8	45B
6106000	19510210	19600518	19.5	45B
6106000	19600518	19690514	9.9	BEND/FRIEZ
6106000	19690514	19700200	10	BEND/FRIEZ
6106000	19700200	99999999	10	U2A/R
6115525	19381100	19500800	12.8	45B
6115525	19500800	19540000	13.4	45B
6115525	19540000	19690600	29.3	U2A
6115525	19690600	99999999	26.9	U2A
6119500	19500818	19590128	11.4	45B
6119500	19590128	19610519	11.8	45B
6119500	19610519	19630110	0	U2A
6119500	19630110	19710222	13.4	U2A
6119500	19710222	19720627	10	U2A
6119500	19720627	19810522	9.9	U2A
6119500	19810522	99999999	10	U2A
6139525	19401207	19520510	15.5	DYNES
6139525	19520510	19571024	18.1	None
6139525	19571024	19620614	0	U2A
6139525	19620614	19700526	10.1	U2A
6139525	19700526	19711100	10	U2A
6139525	19711100	99999999	10	U2A/R
6144475	19400718	19591125	12.8	45B
6144475	19591125	19650222	12.2	45B
6144475	19650222	19670210	12.5	45B

Continued on next page

Table B.2 – continued from previous page

Station ID	Start (yyyymmdd)	End (yyyymmdd)	Hgt. (m)	Anemometer Type
6144475	19670210	19700217	9.9	45B
6144475	19670210	19770300	10	U2A
6144475	19700217	19790629	10	45B
6144475	19770300	99999999	10	U2A/R
6158665	19620803	19690826	17.9	45B
6158665	19650106	19690826	11.4	U2A
6158665	19690826	99999999	11.4	U2A/R
6158733	19580415	19640529	16.7	45B
6158733	19640529	19660225	10.3	BEND/FRIEZ
6158733	19640529	19730700	10.1	U2A
6158733	19660225	19680308	10.2	BEND/FRIEZ
6158733	19680308	19790613	10.3	BEND/FRIEZ
6158733	19730700	19760400	10.1	U2A/R
6158733	19790613	99999999	10	U2A/R
6158875	19410807	19450924	0	DYNES
6158875	19450924	19590417	18.5	DYNES
6158875	19590417	19621004	22.8	U2A
6158875	19621004	19650115	9.1	U2A
6158875	19650115	19700400	10	U2A
6158875	19700400	99999999	10	U2A/R

Appendix C

Atlas of Wind Predictability

C.1 Cross-validated r^2 predictability plots

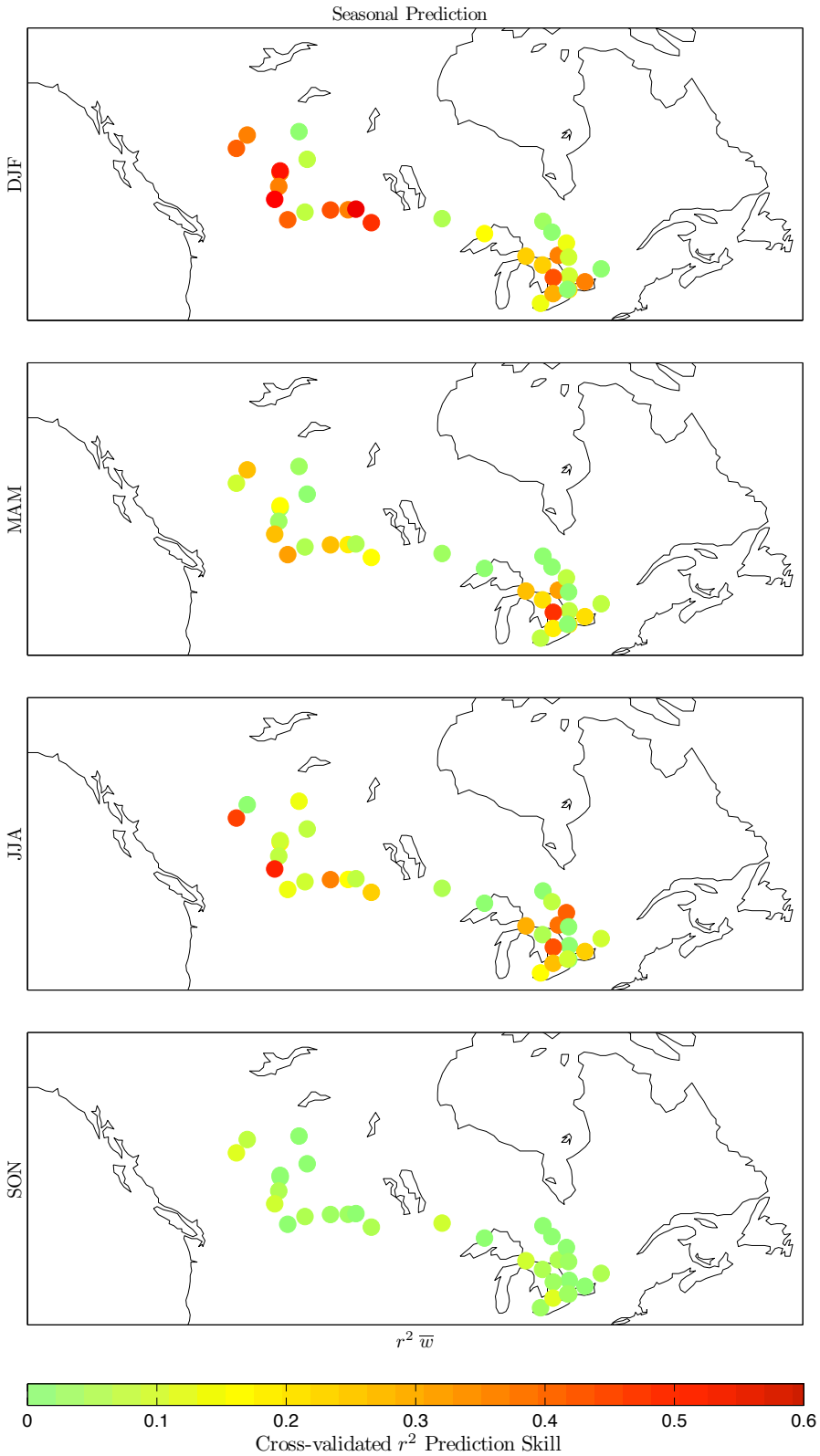

Figure C.1: As with Figure 3.8 for seasonally-averaged $\overline{w}$.

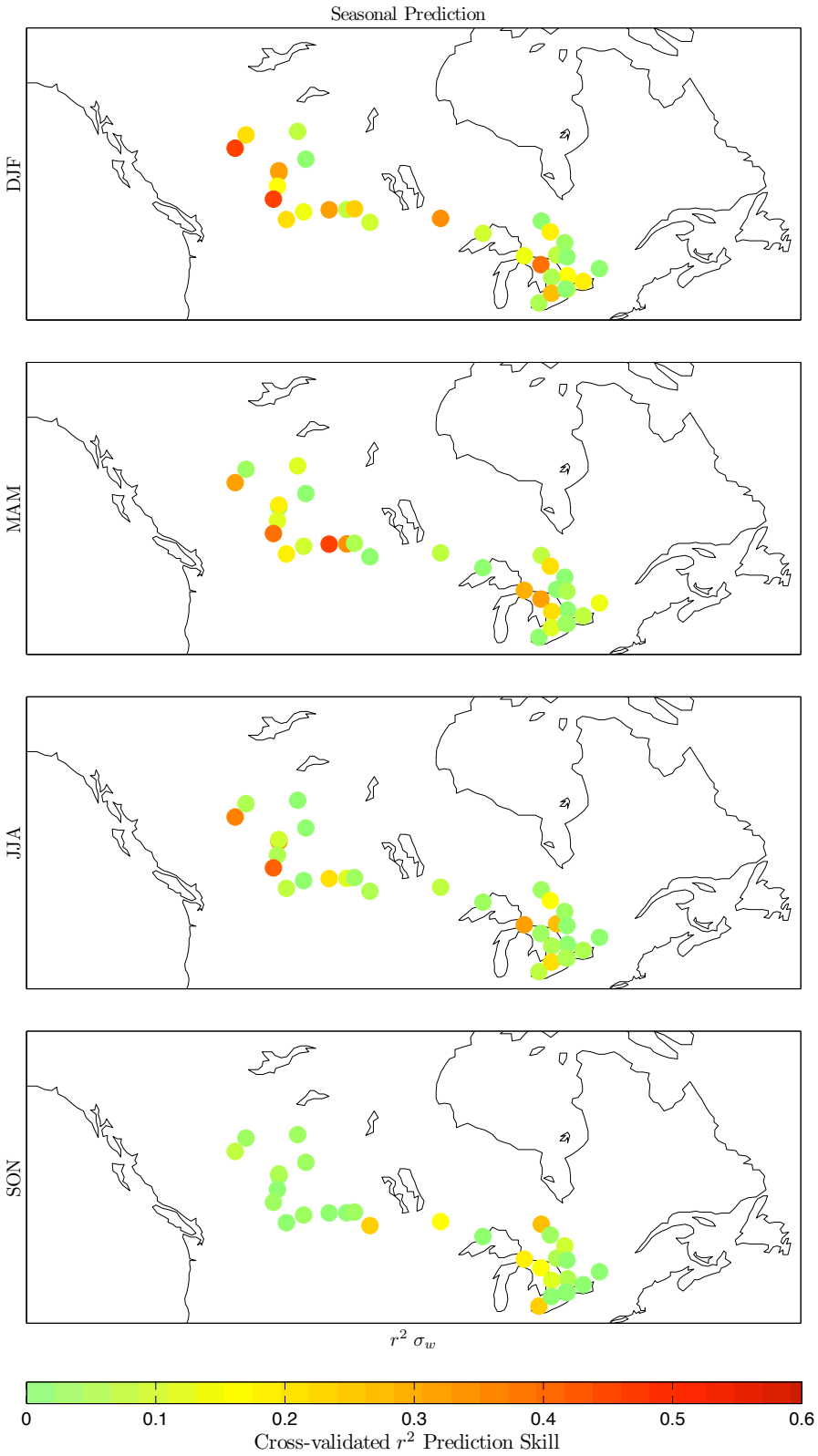

Figure C.2: As with Figure 3.8 for seasonally-averaged σ_w.

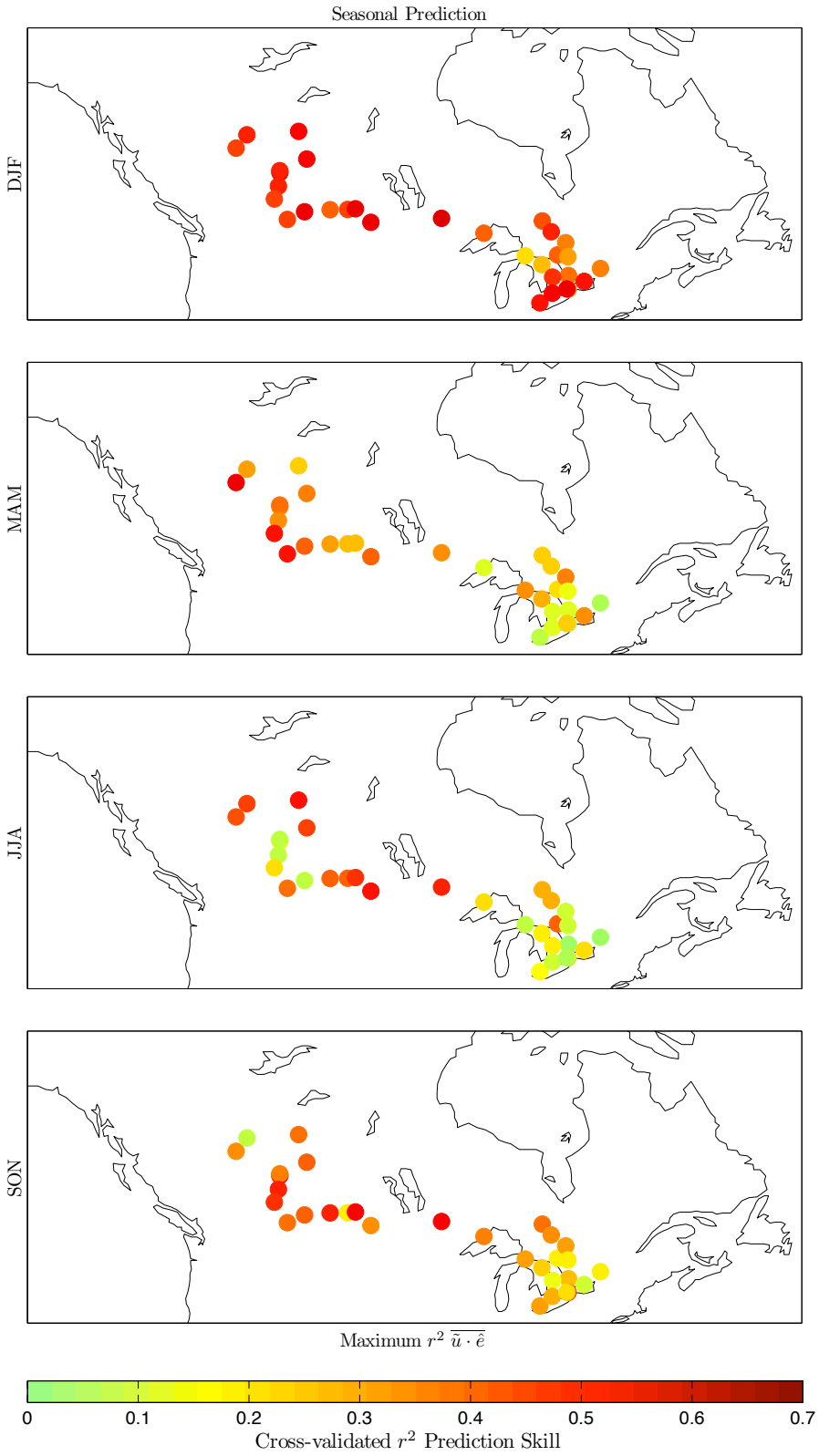

Figure C.3: As with Figure 3.8 for seasonally-averaged $\overline{\tilde{u} \cdot \hat{e}}$.

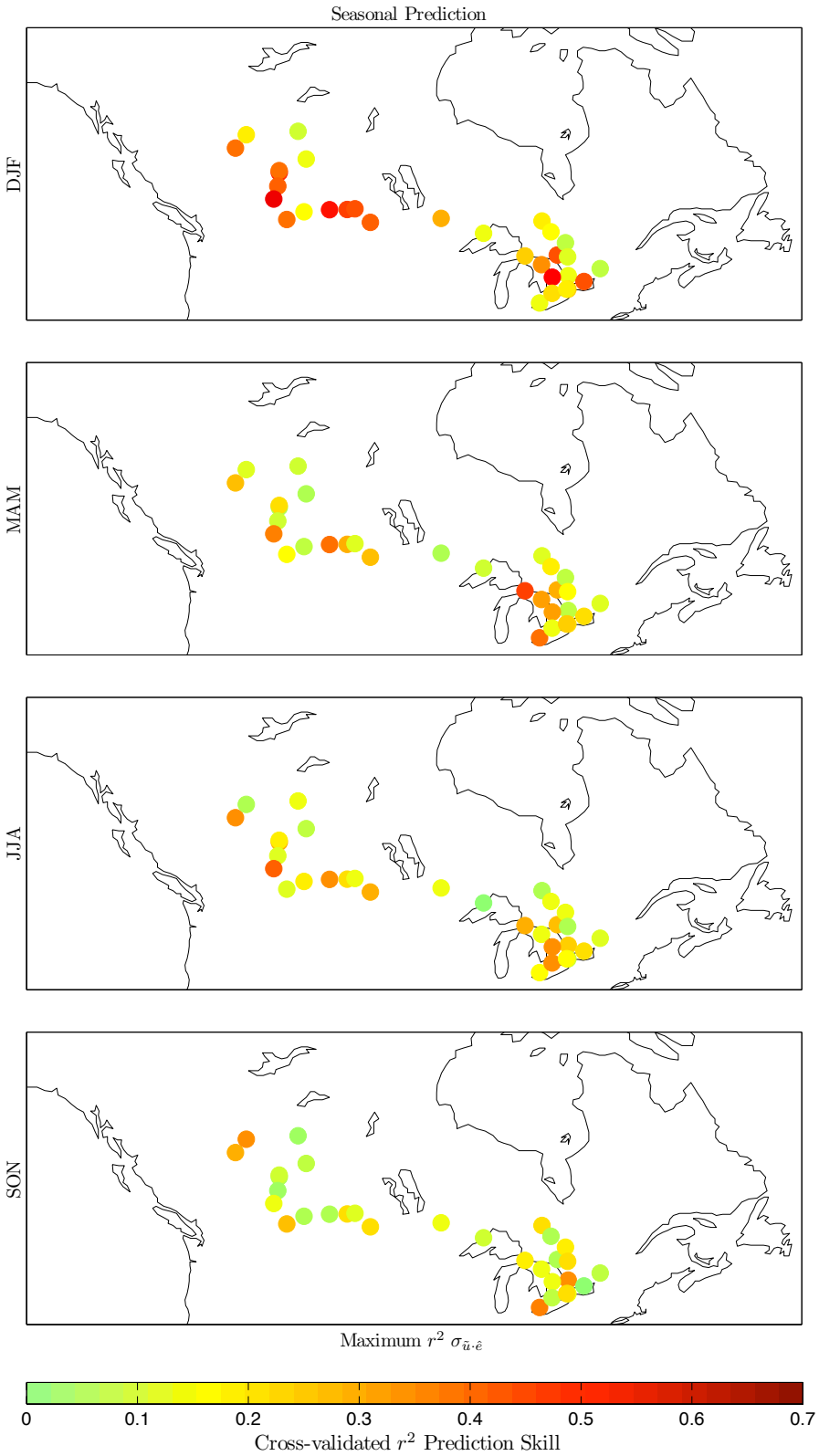

Figure C.4: As with Figure 3.8 for seasonally-averaged $\sigma_{\tilde{u}\cdot\hat{e}}$.

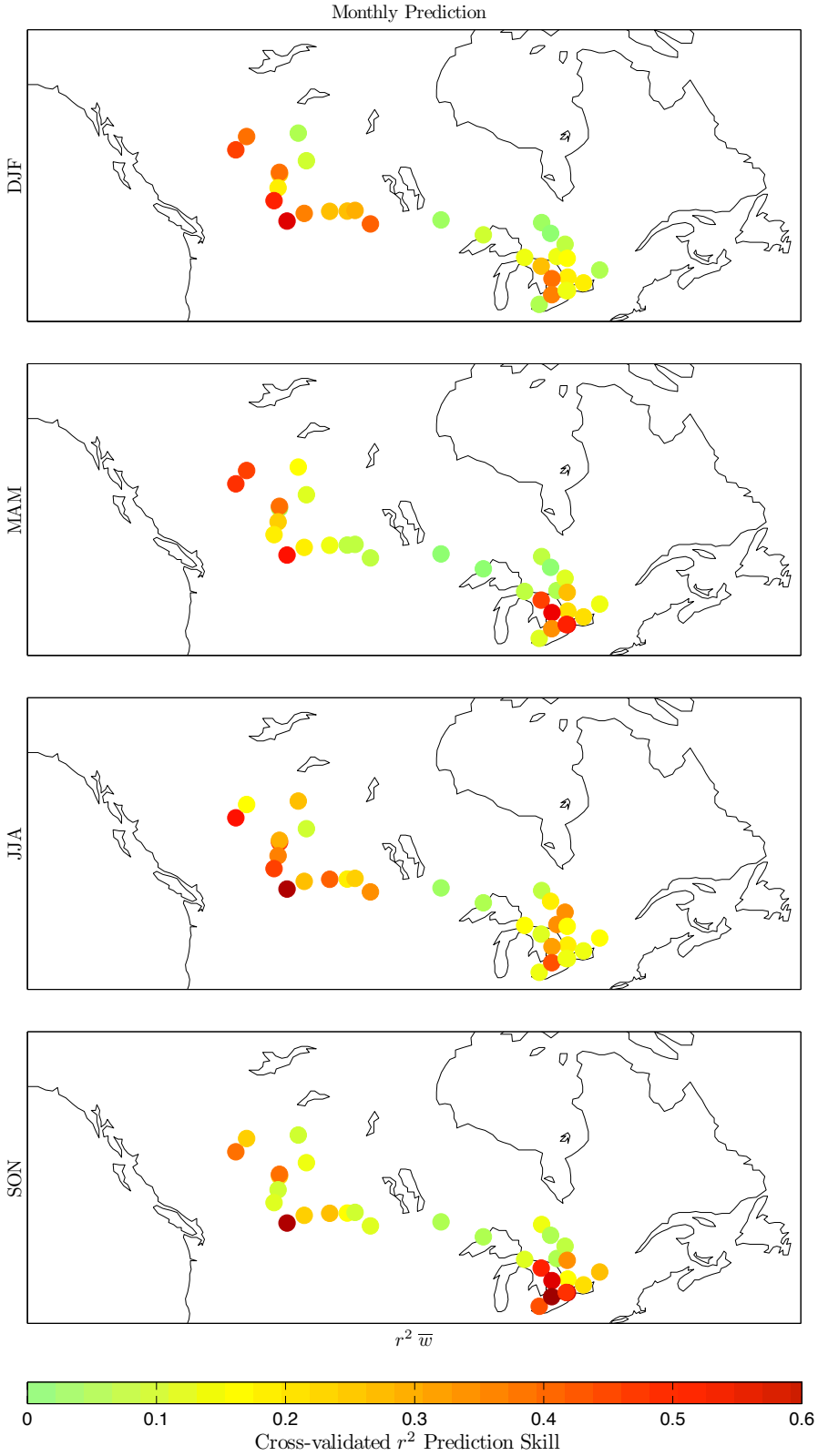

Figure C.5: As with Figure 3.8 for monthly-averaged $\overline{w}$.

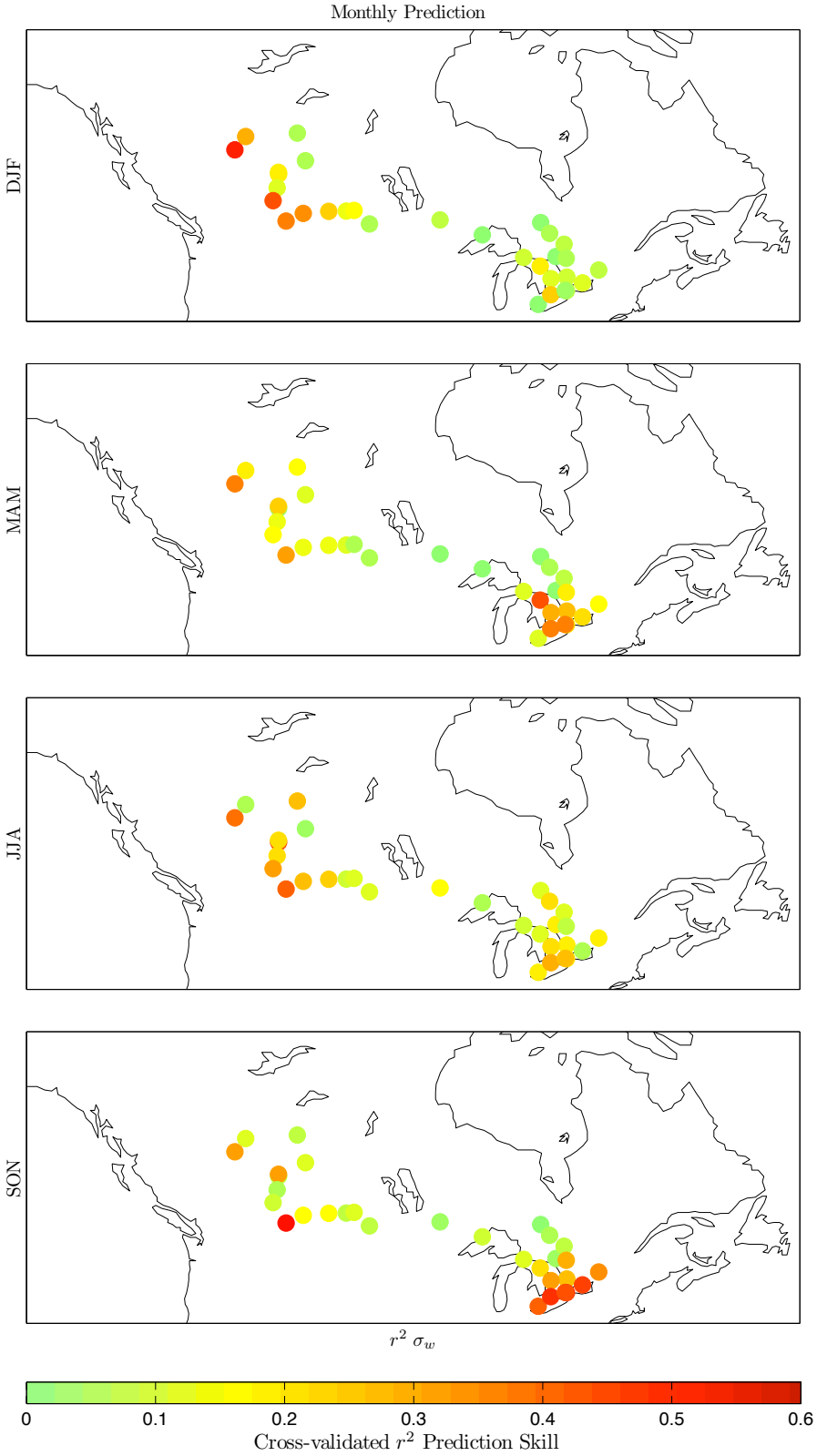

Figure C.6: As with Figure 3.8 for monthly-averaged σ_w.

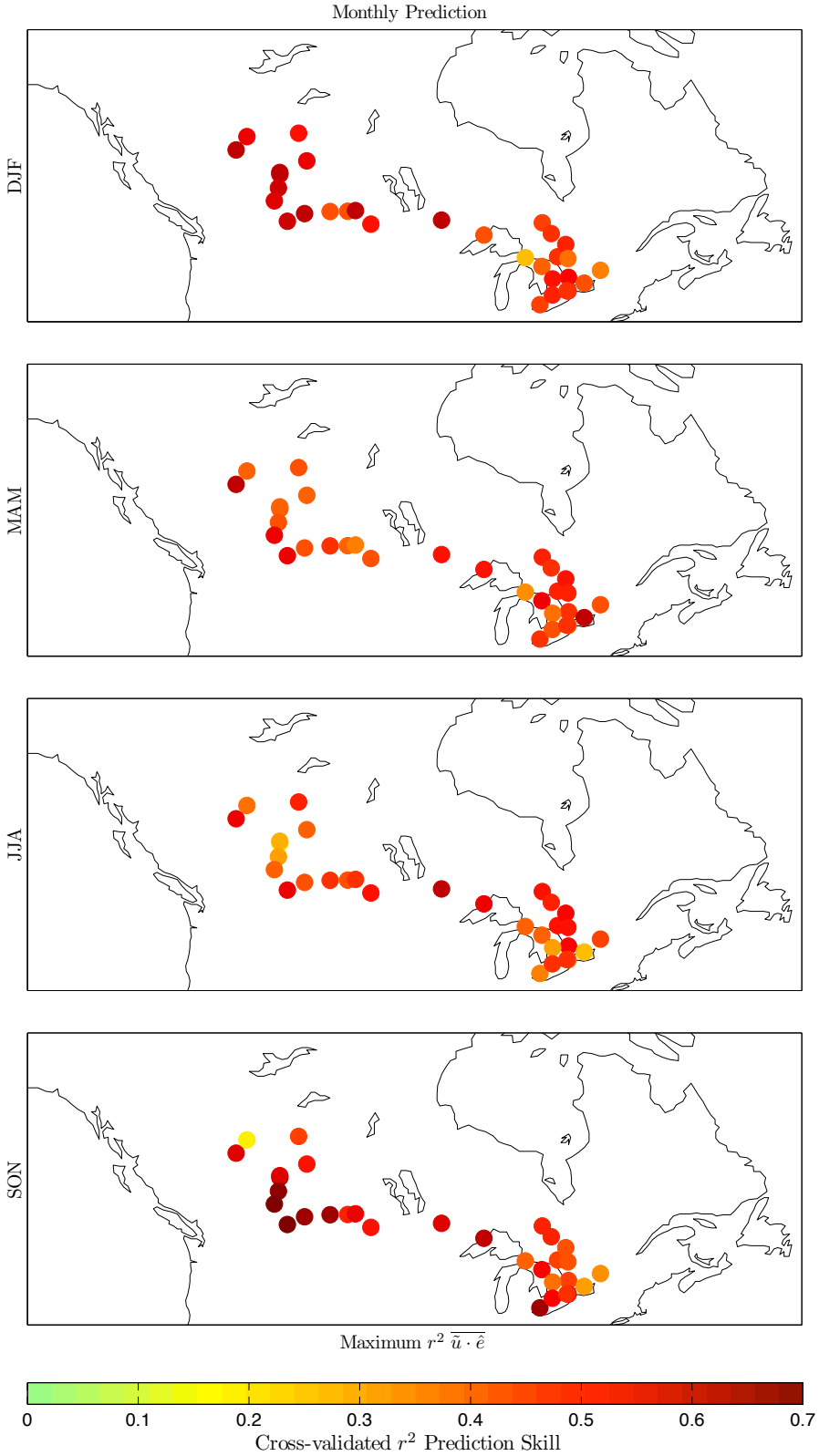

Figure C.7: As with Figure 3.8 for monthly-averaged $\widetilde{u} \cdot \hat{e}$.

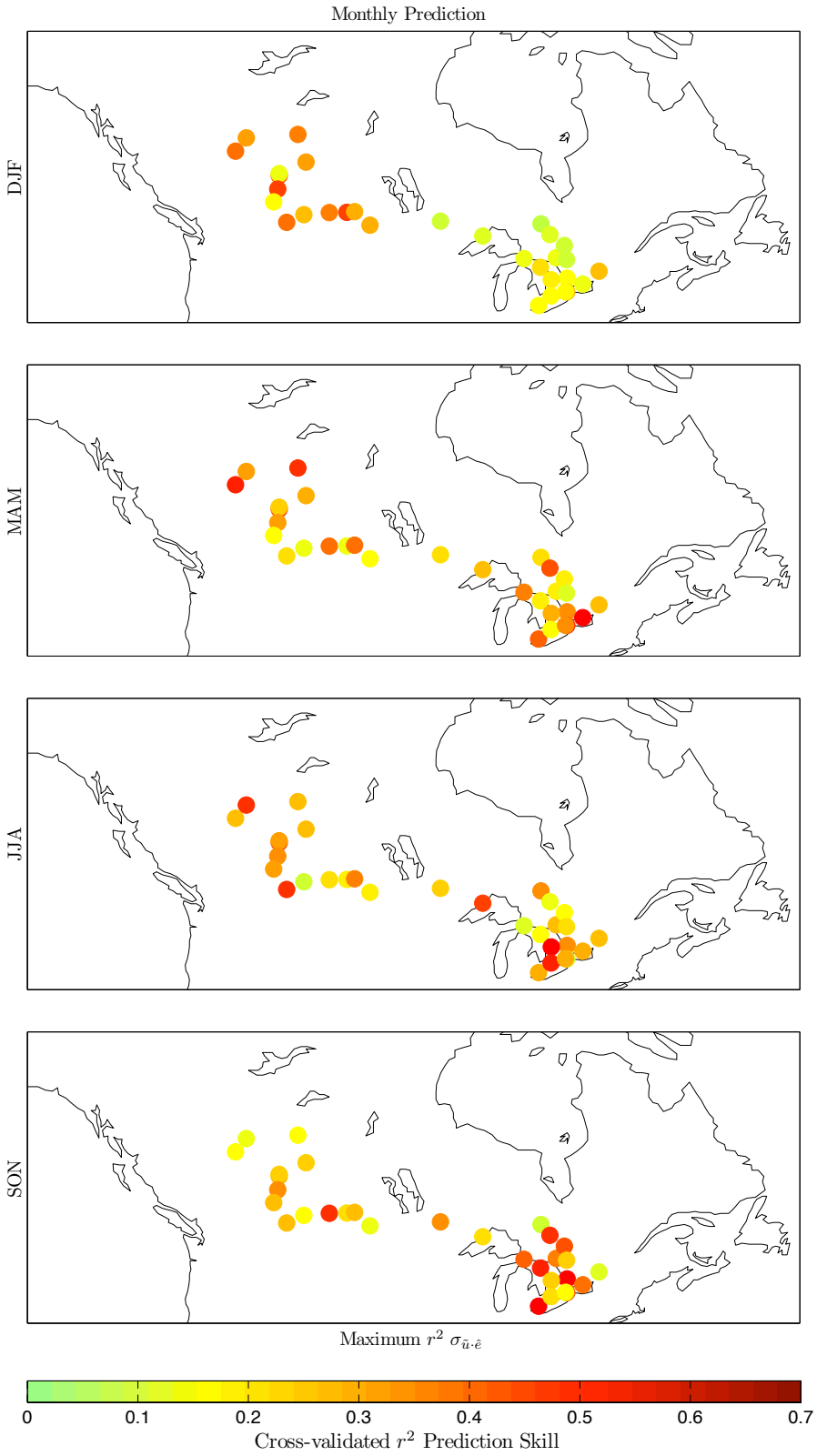

Figure C.8: As with Figure 3.8 for monthly-averaged $\sigma_{\tilde{u}\cdot\hat{e}}$.

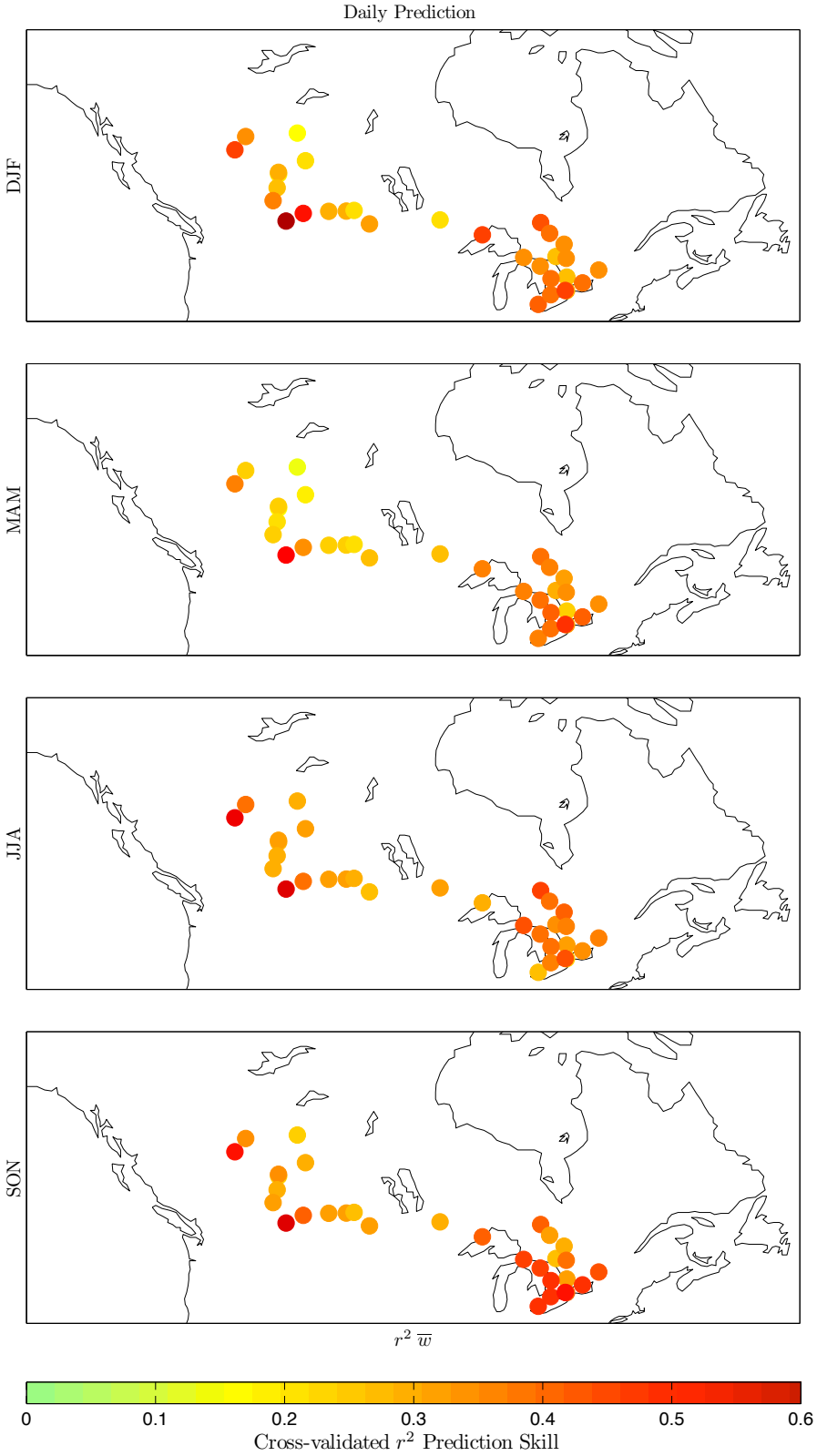

Figure C.9: As with Figure 3.8 for daily-averaged $\overline{w}$.

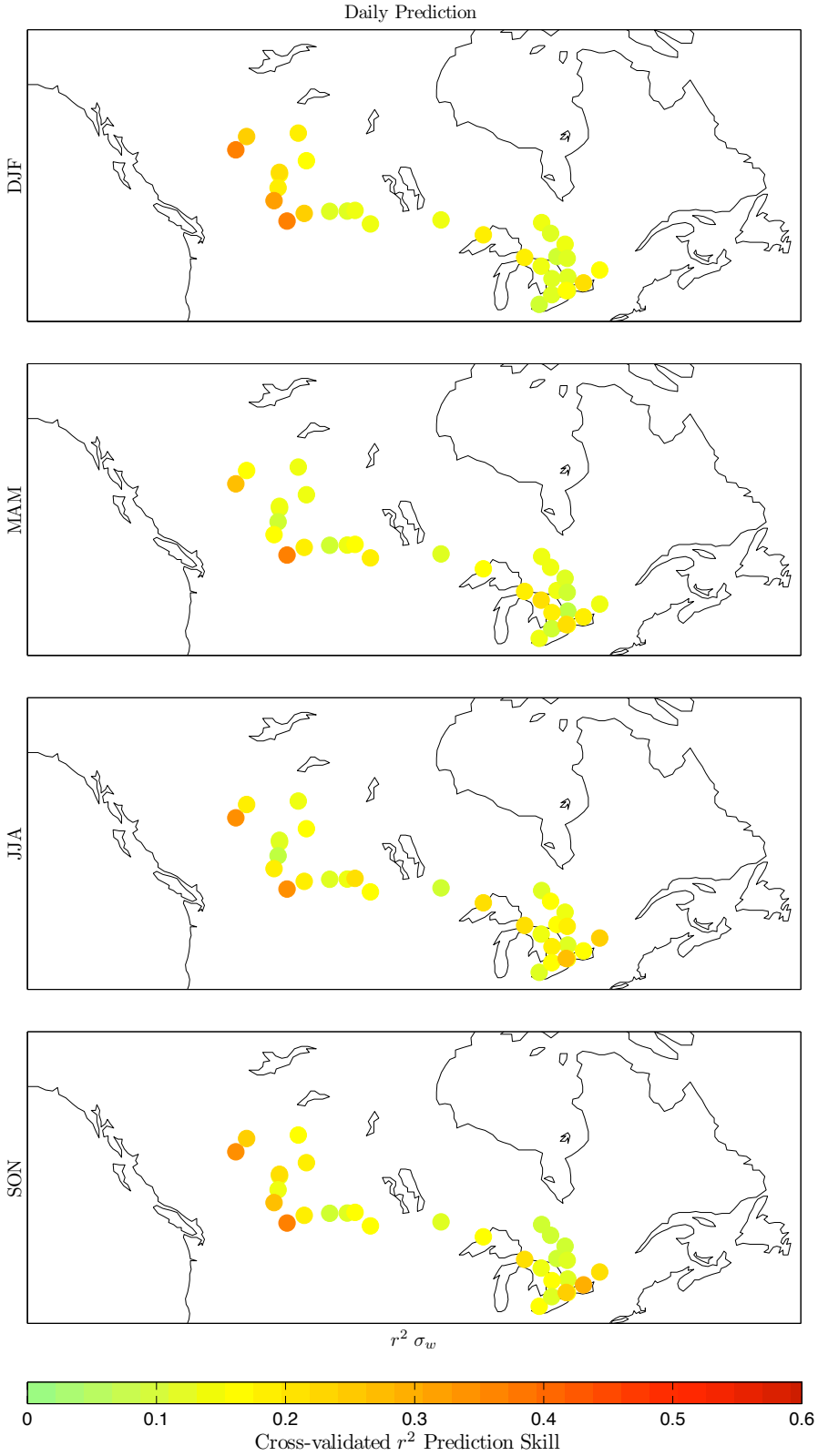

Figure C.10: As with Figure 3.8 for daily-averaged σ_w.

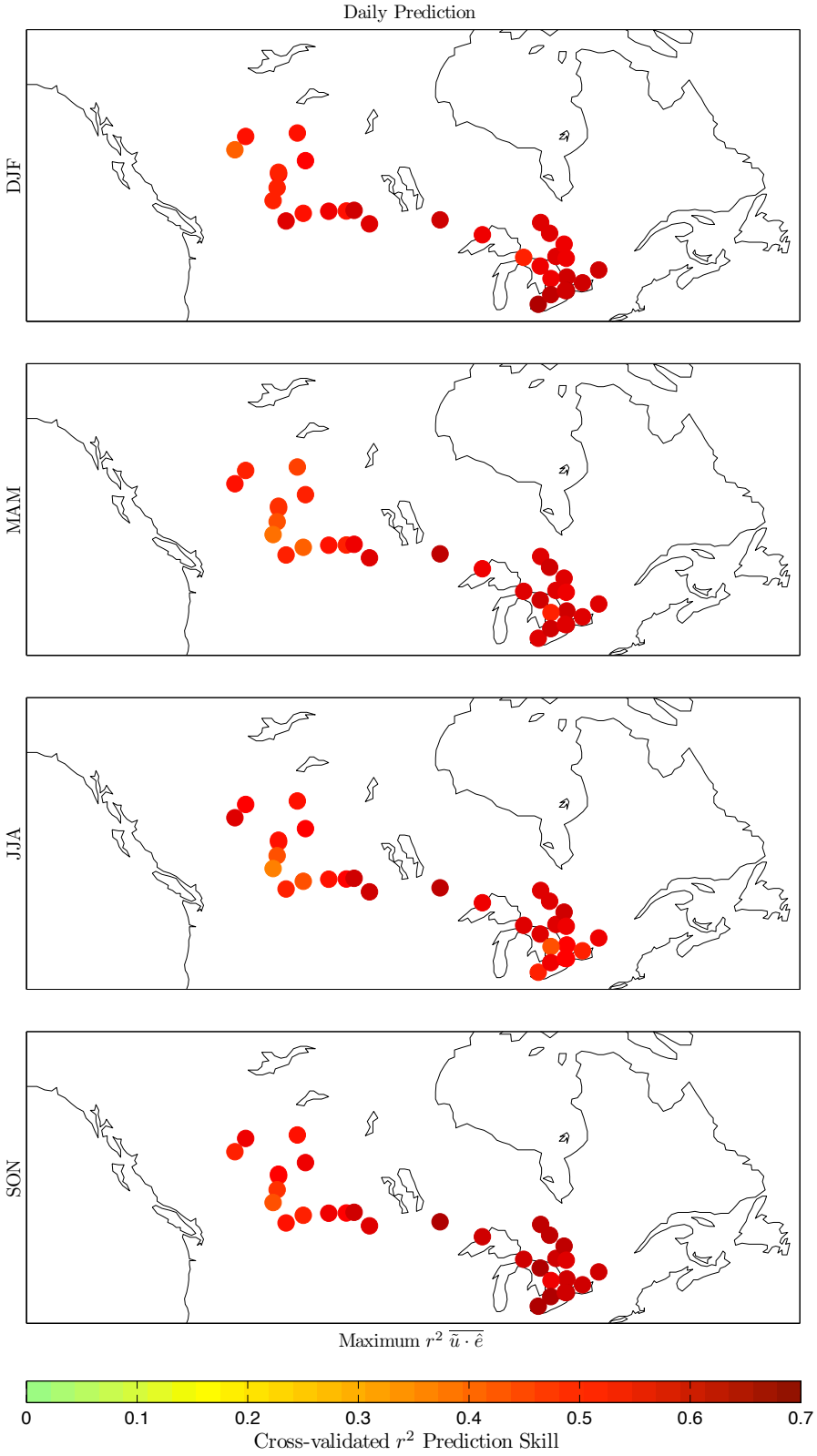

Figure C.11: As with Figure 3.8 for daily-averaged $\overline{\tilde{u} \cdot \hat{e}}$.

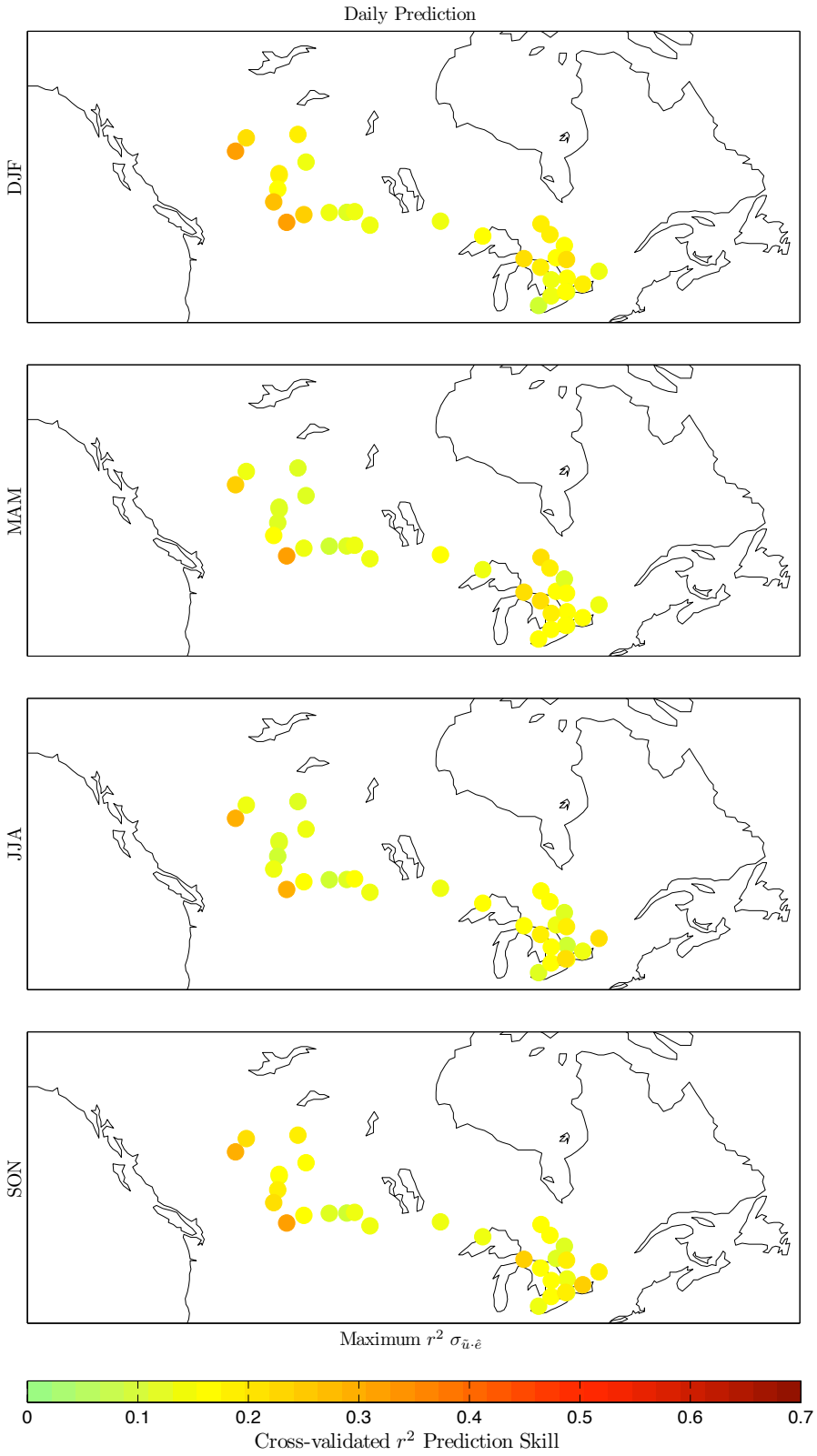

Figure C.12: As with Figure 3.8 for daily-averaged $\sigma_{\tilde{u}\cdot\hat{e}}$.

C.2 Polar Prediction Skill Plots

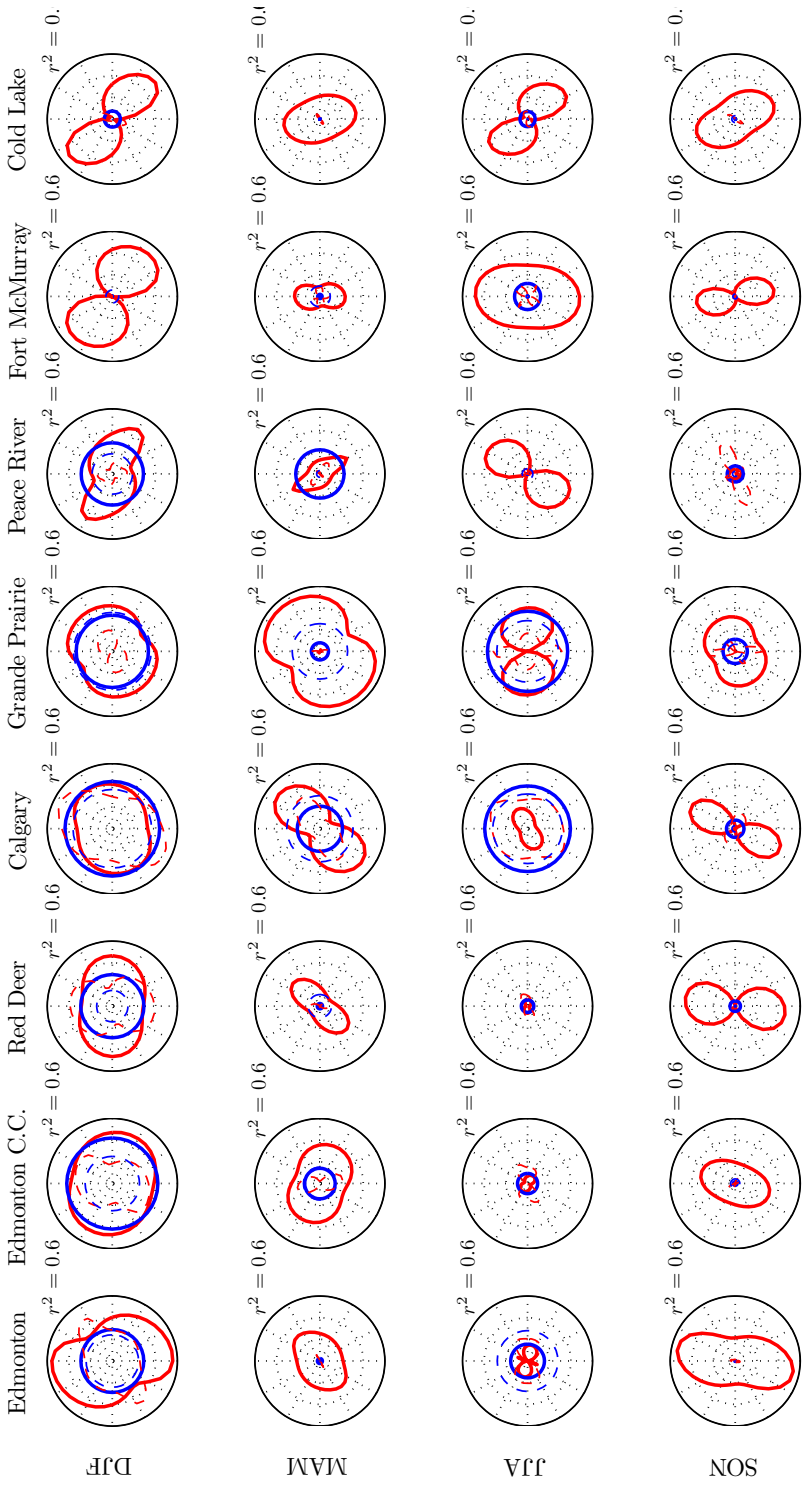

Figure C.13: Cross-validated r^2 prediction skill of seasonally-averaged wind quantities (across northern plains sites).

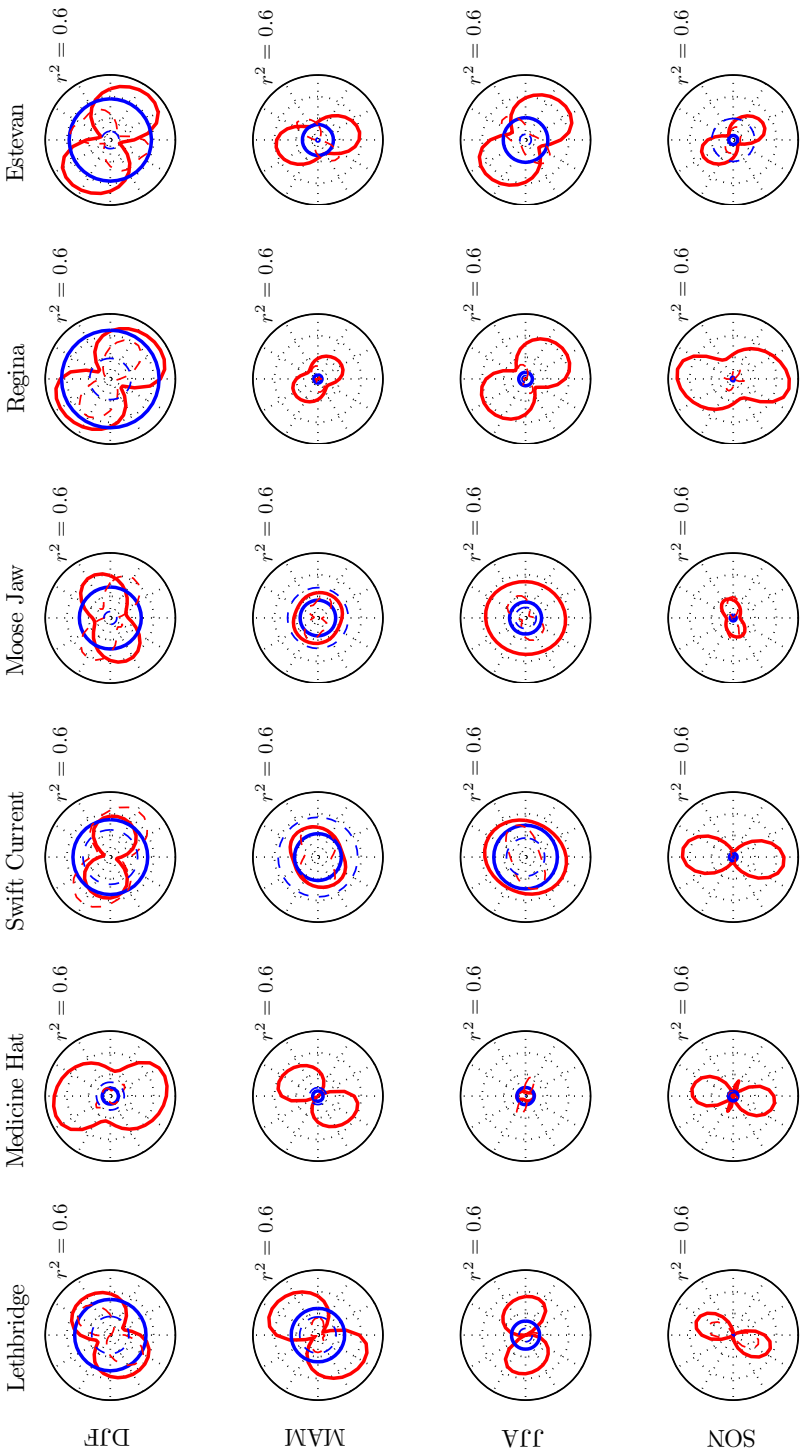

Figure C.14: Cross-validated r^2 prediction skill of seasonally-averaged wind quantities (across southern prairies sites).

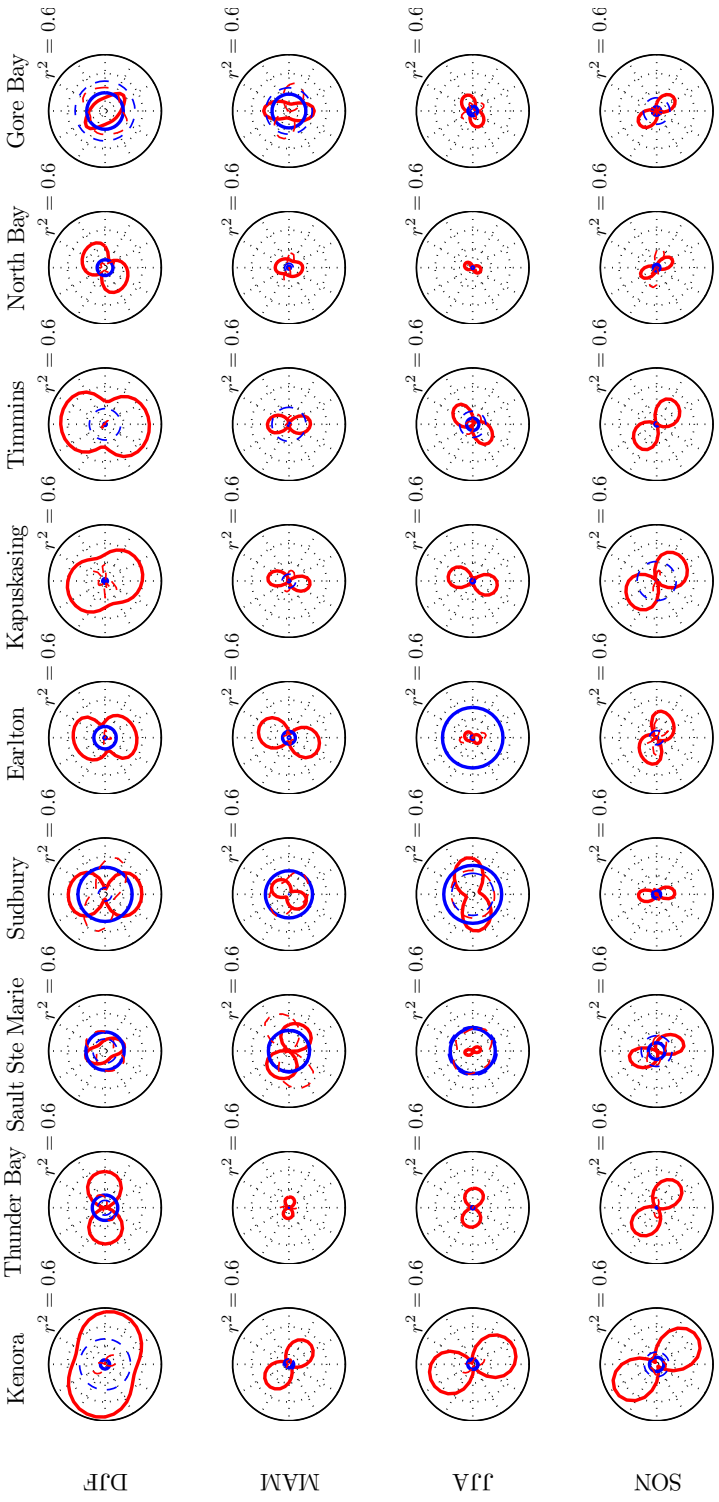

Figure C.15: Cross-validated r^2 prediction skill of seasonally-averaged wind quantities (across Ontario sites - 1 of 2).

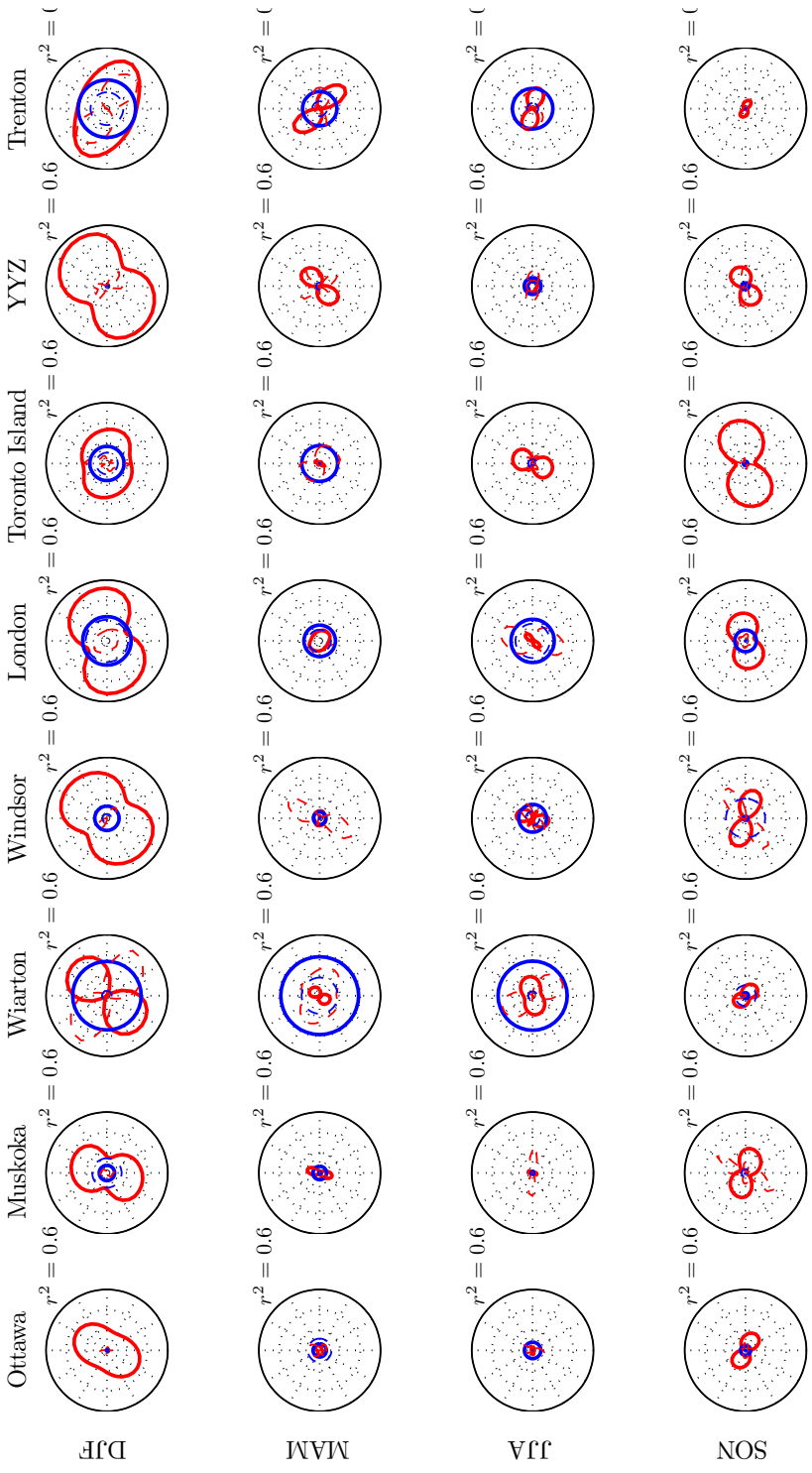

Figure C.16: Cross-validated r^2 prediction skill of seasonally-averaged wind quantities (across Ontario sites - 2 of 2).

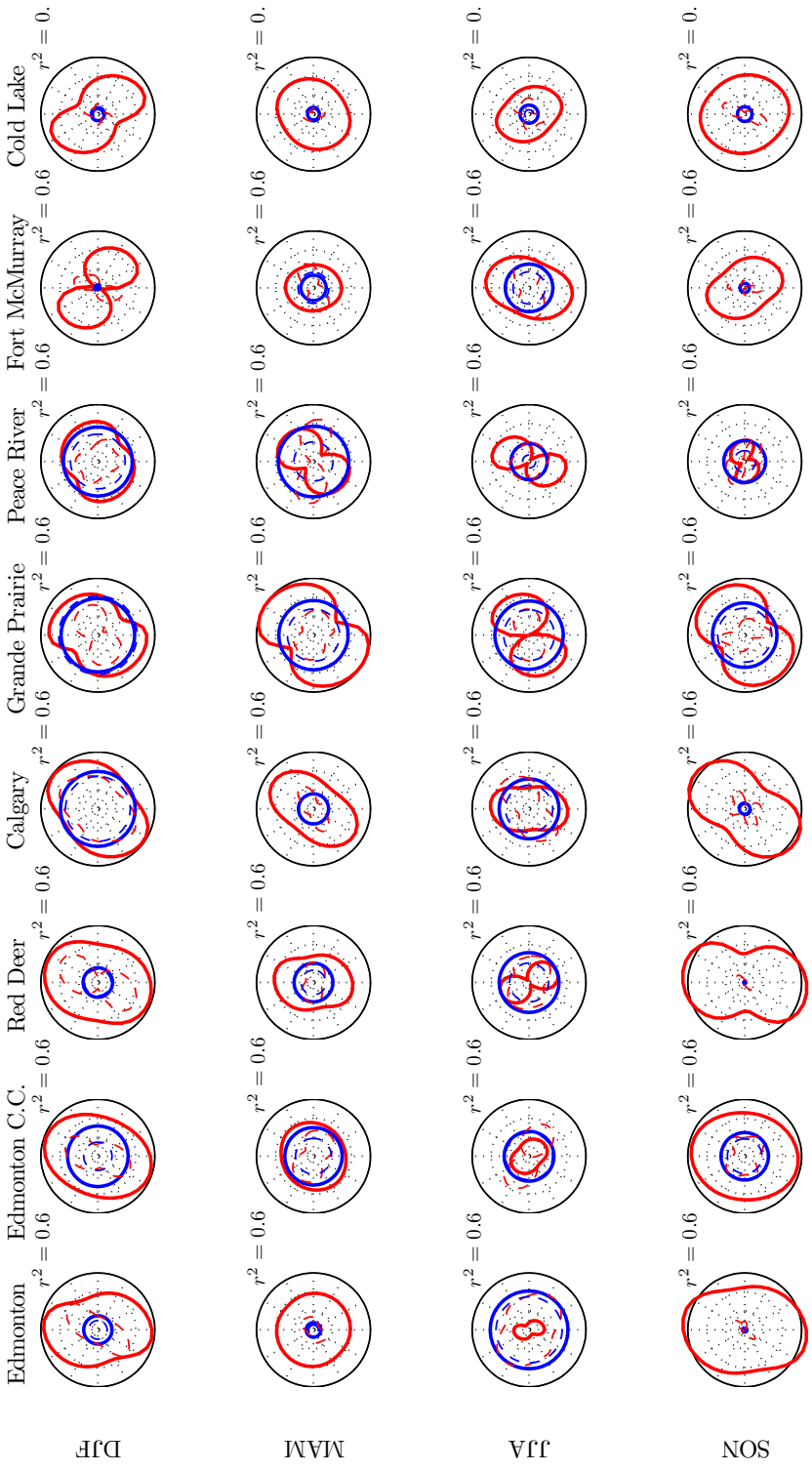

Figure C.17: Cross-validated r^2 prediction skill of monthly-averaged wind quantities (across northern plains sites).

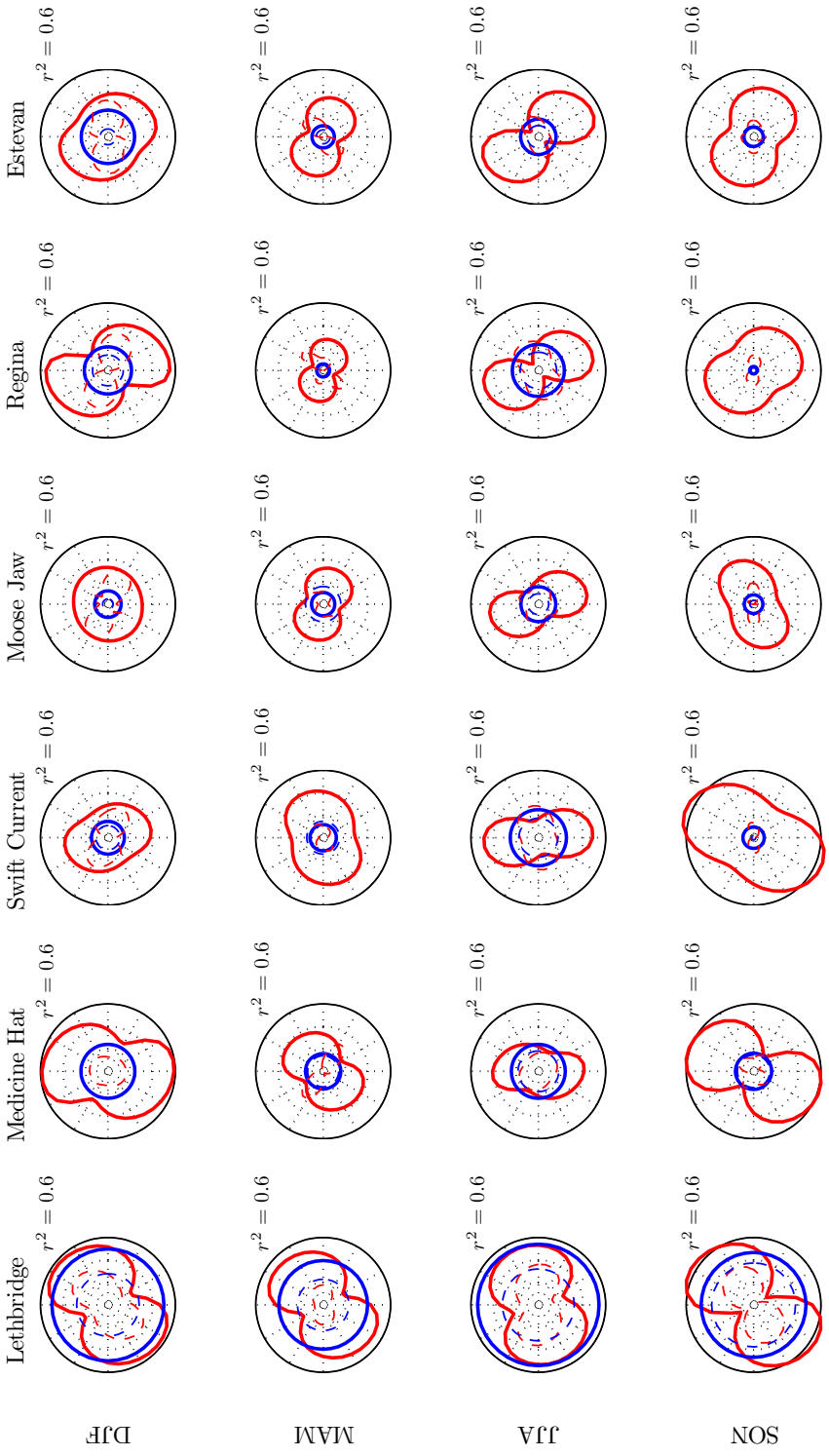

Figure C.18: Cross-validated r^2 prediction skill of monthly-averaged wind quantities (across southern prairies sites).

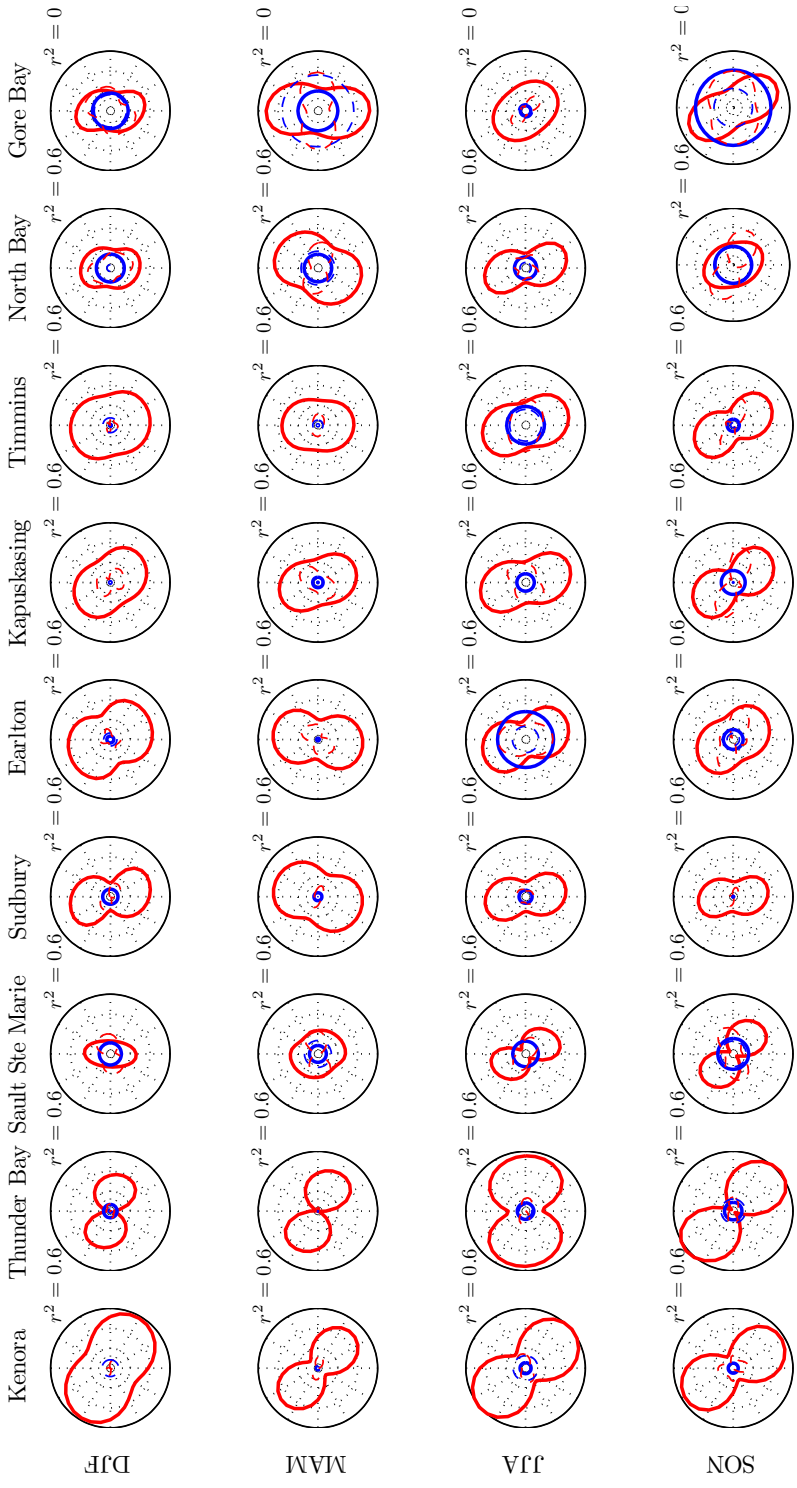

Figure C.19: Cross-validated r^2 prediction skill of monthly-averaged wind quantities (across Ontario sites - 1 of 2).

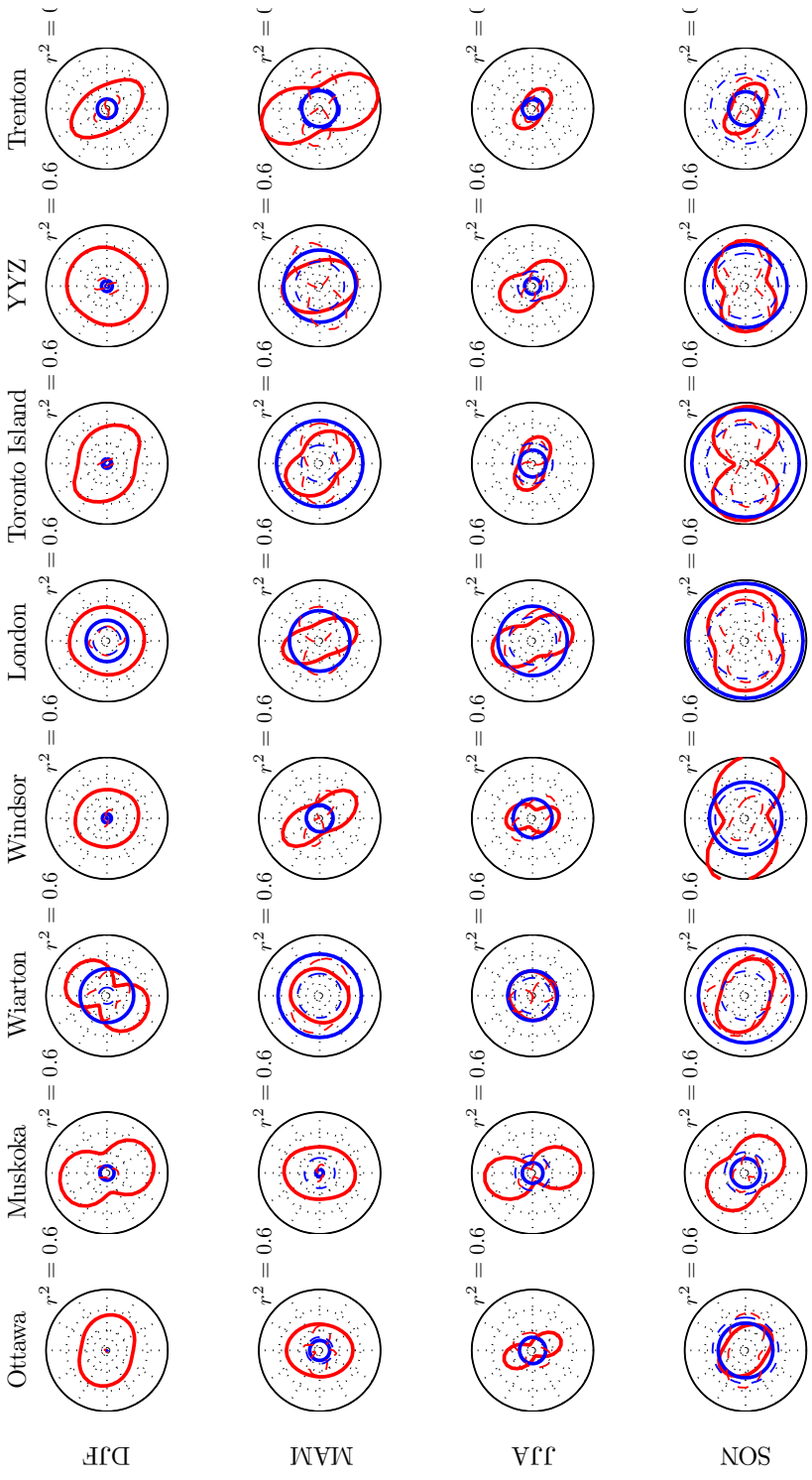

Figure C.20: Cross-validated r^2 prediction skill of monthly-averaged wind quantities (across Ontario sites - 2 of 2).

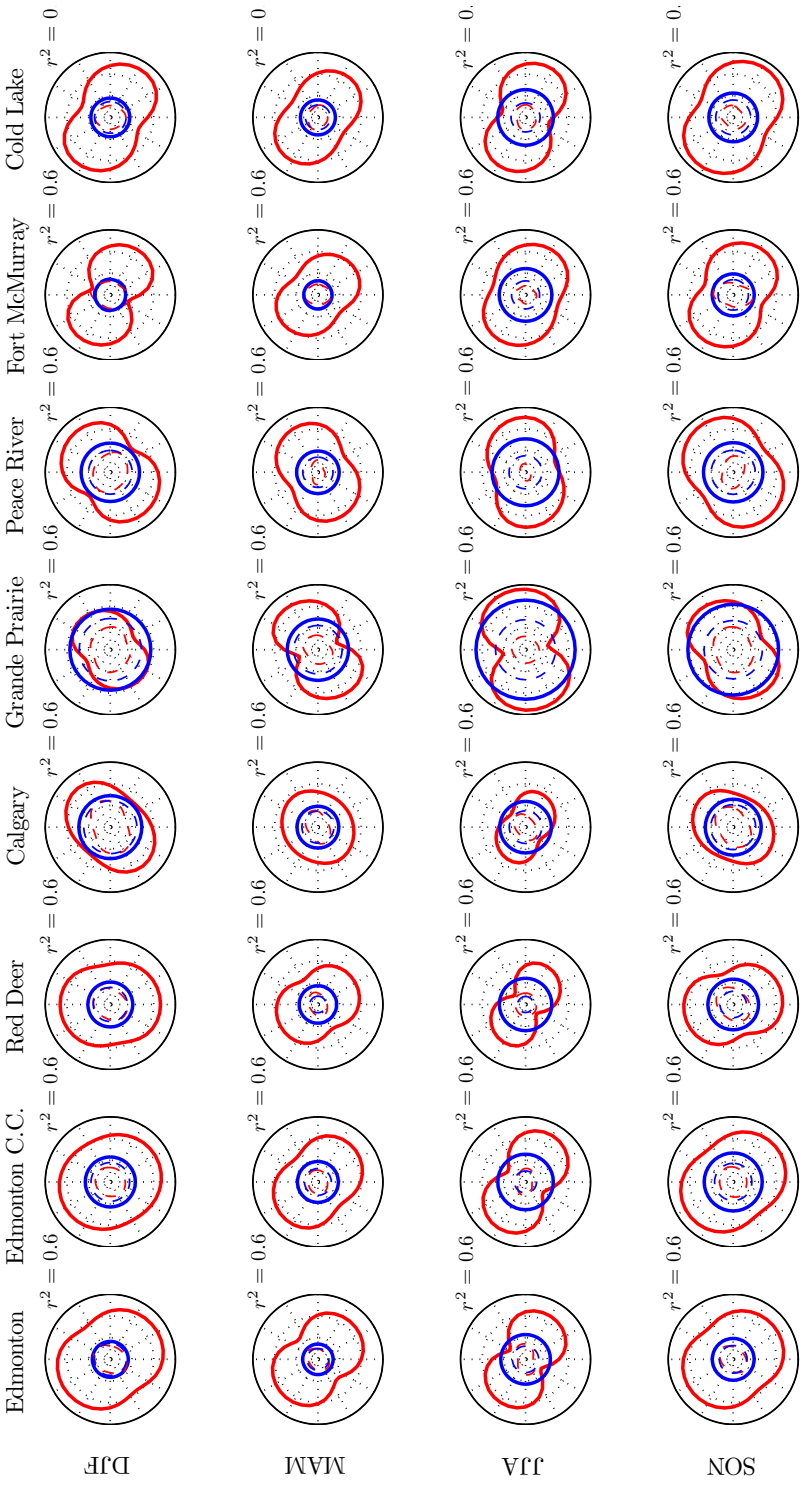

Figure C.21: Cross-validated r^2 prediction skill of daily-averaged wind quantities (across northern plains sites).

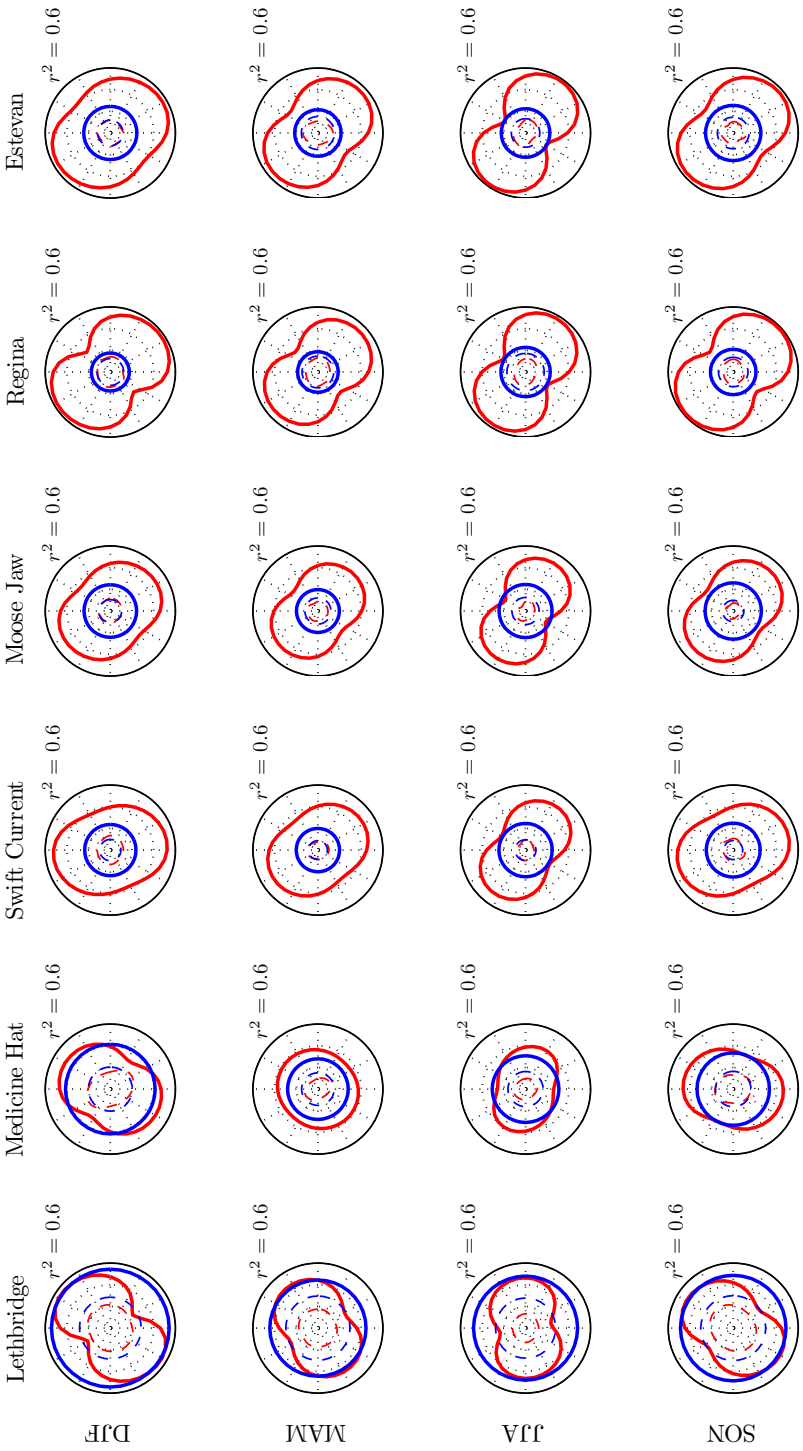

Figure C.22: Cross-validated r^2 prediction skill of daily-averaged wind quantities (across southern prairies sites).

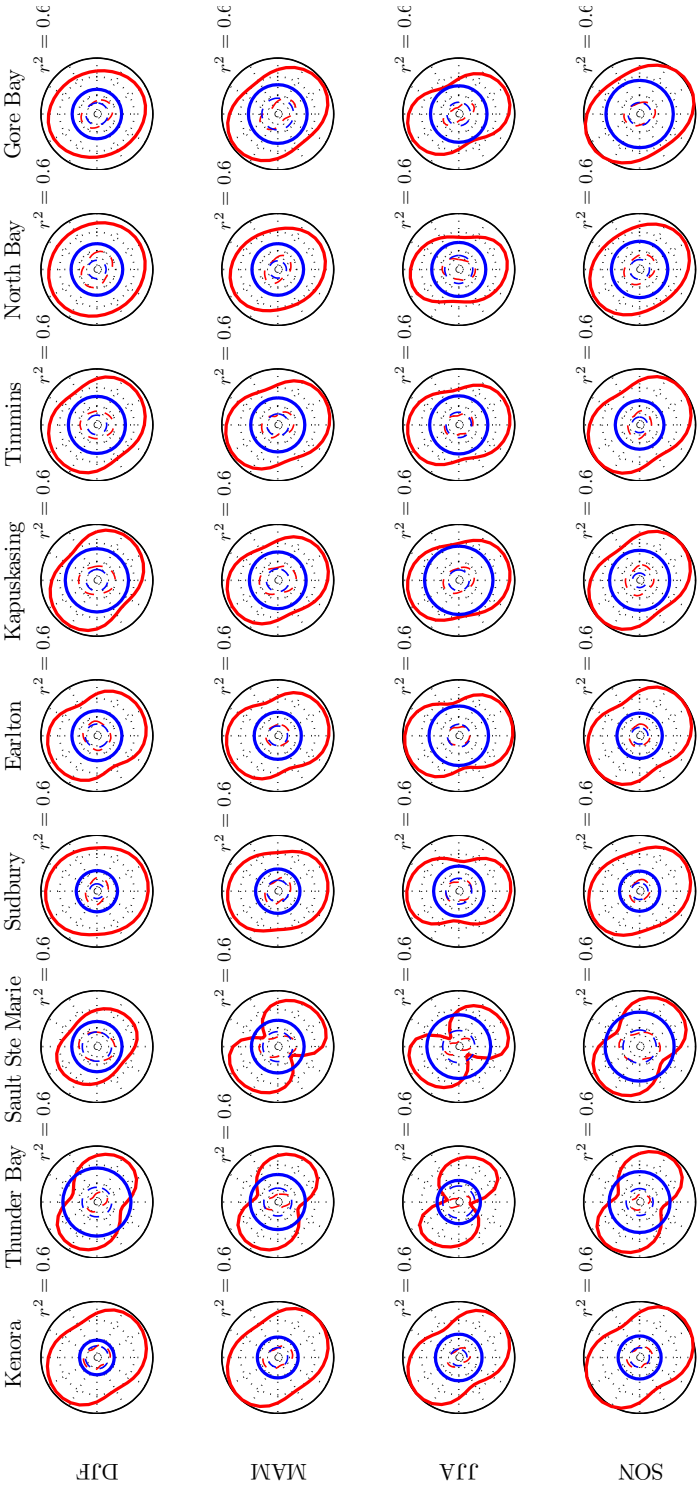

Figure C.23: Cross-validated r^2 prediction skill of daily-averaged wind quantities (across Ontario sites - 1 of 2).

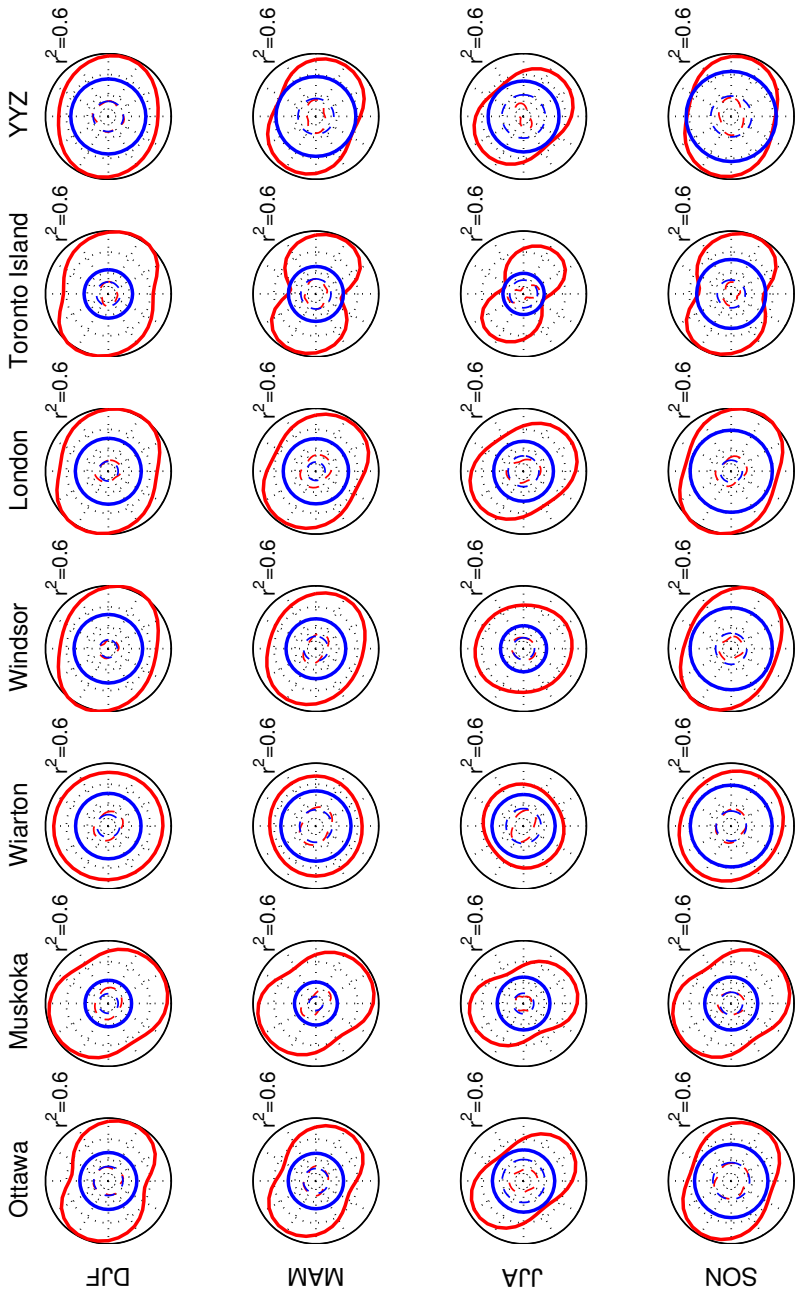

Figure C.24: Cross-validated r^2 prediction skill of daily-averaged wind quantities (across Ontario sites - 2 of 2).

C.3 Polar Correlation Plots

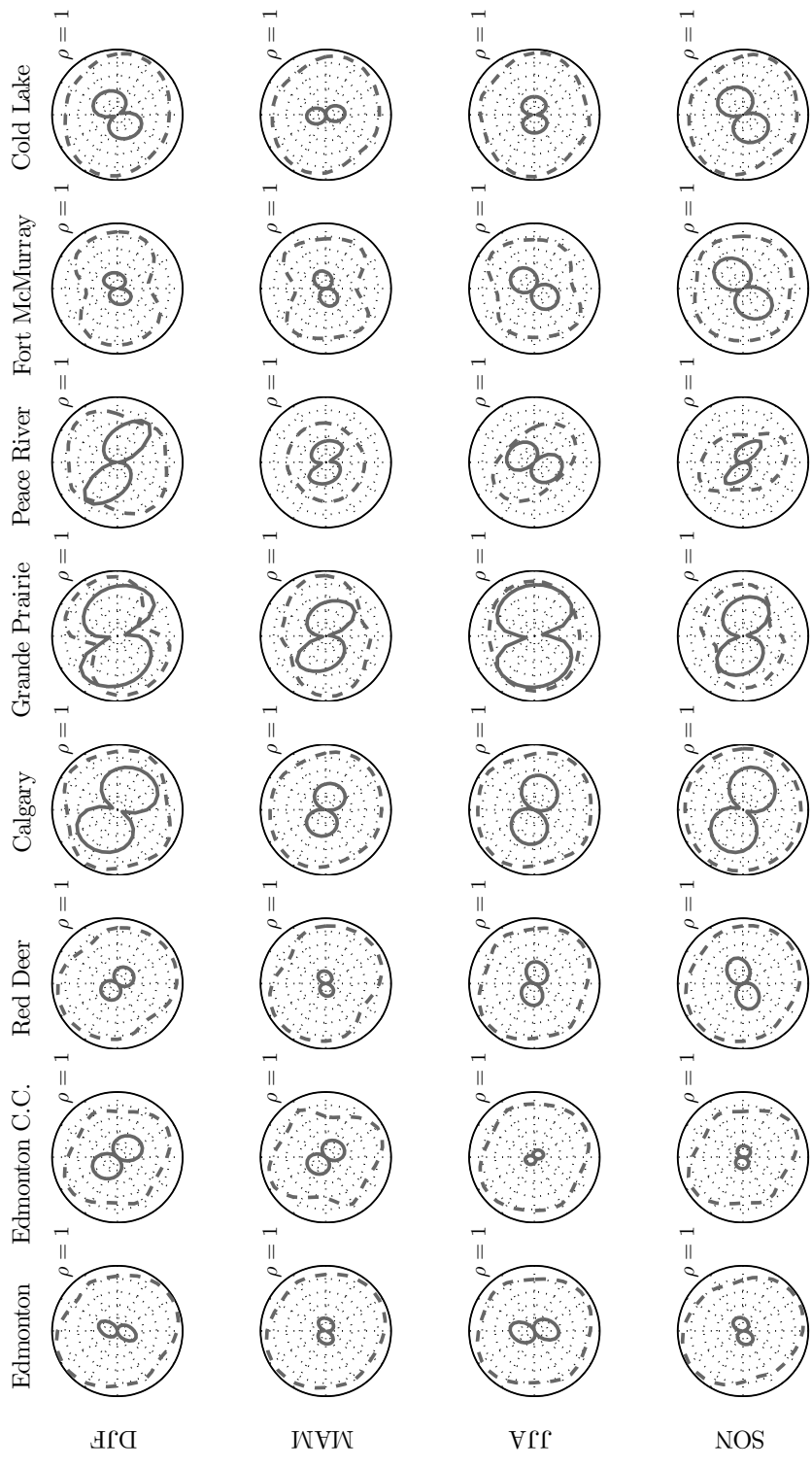

Figure C.25: As with Figure 4.11 for seasonally-averaged wind quantities (across northern plains sites).

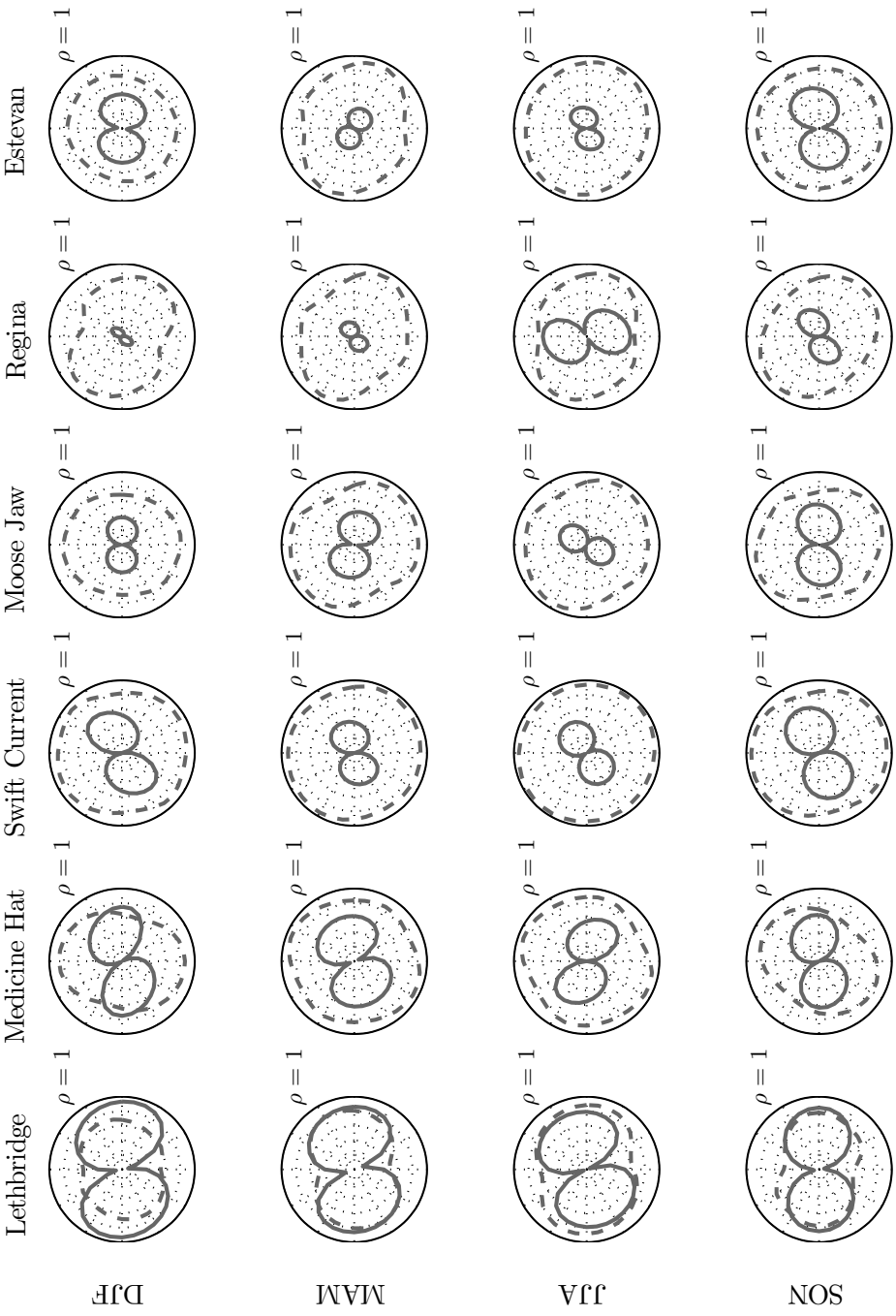

Figure C.26: As with Figure 4.11 for seasonally-averaged wind quantities (across southern prairies sites).

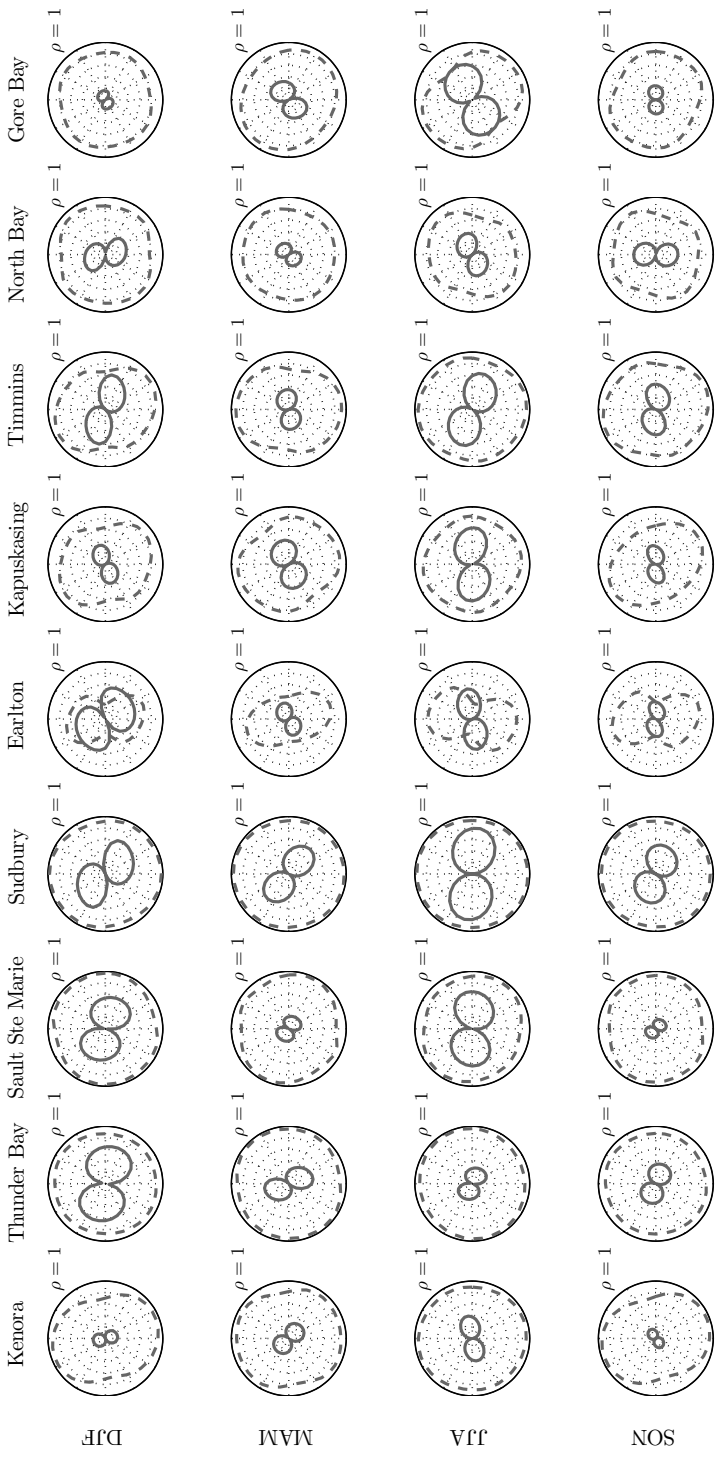

Figure C.27: As with Figure 4.11 for seasonally-averaged wind quantities (across Ontario sites - 1 of 2).

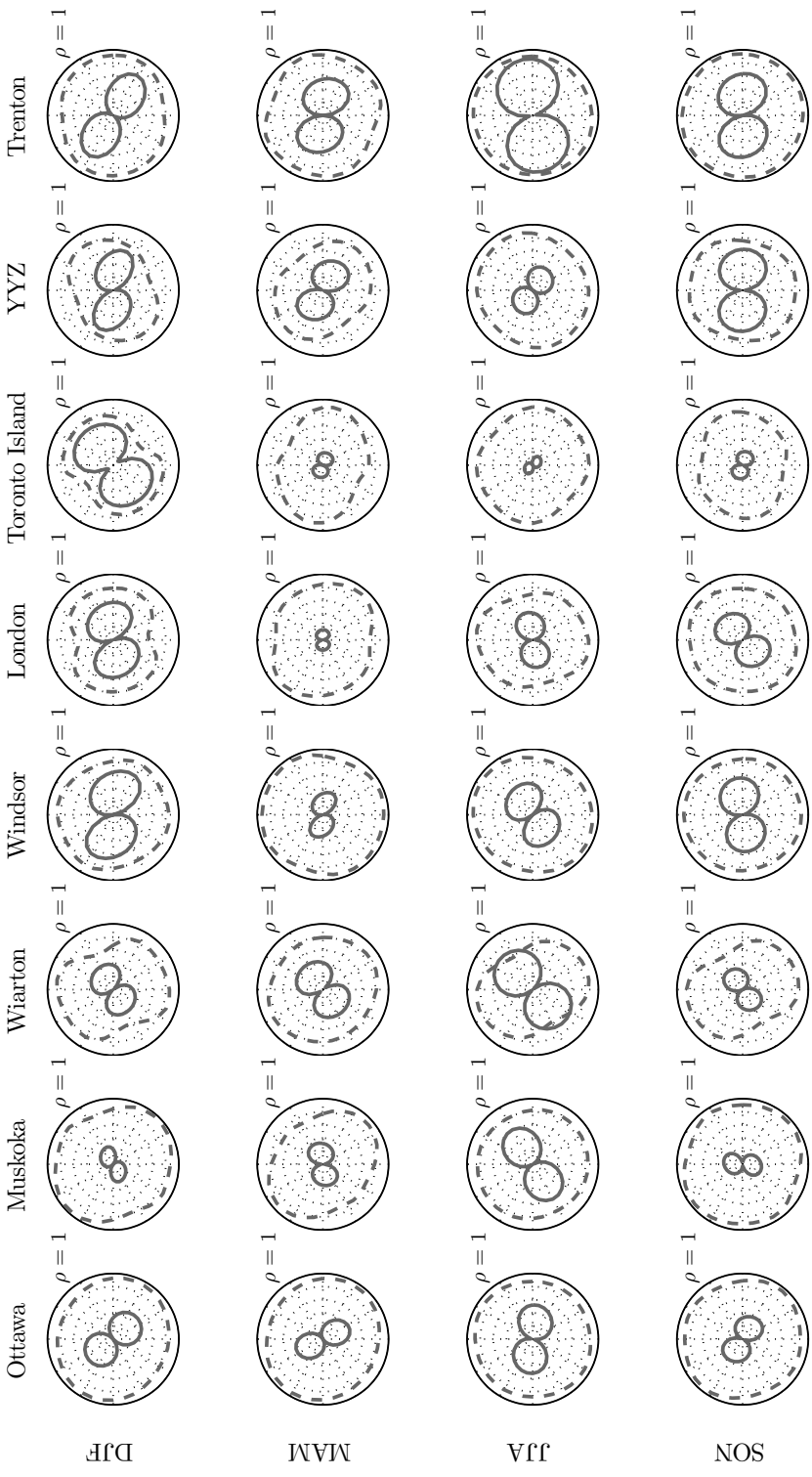

Figure C.28: As with Figure 4.11 for seasonally-averaged wind quantities (across Ontario sites - 2 of 2).

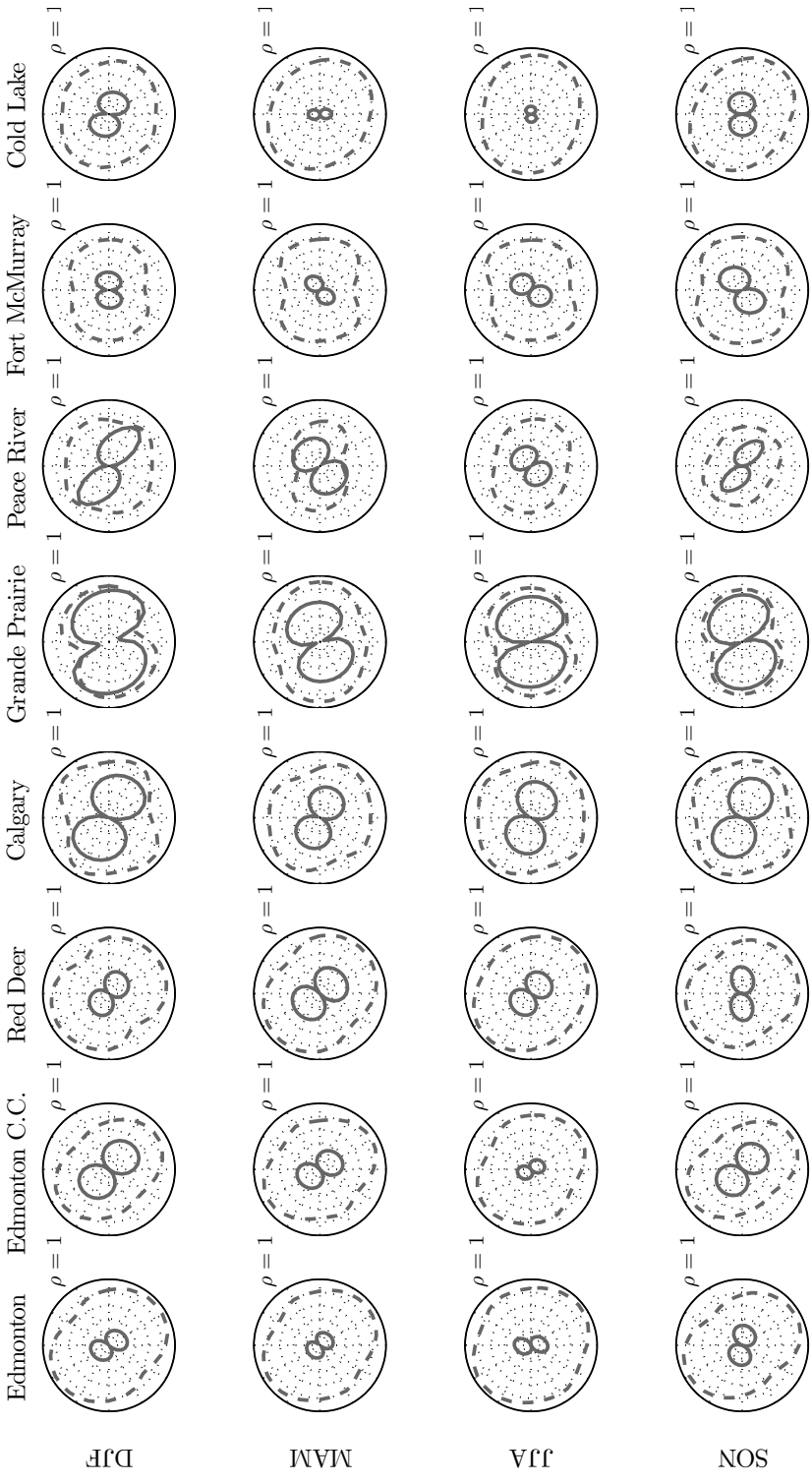

Figure C.29: As with Figure 4.11 for monthly-averaged wind quantities (across northern plains sites).

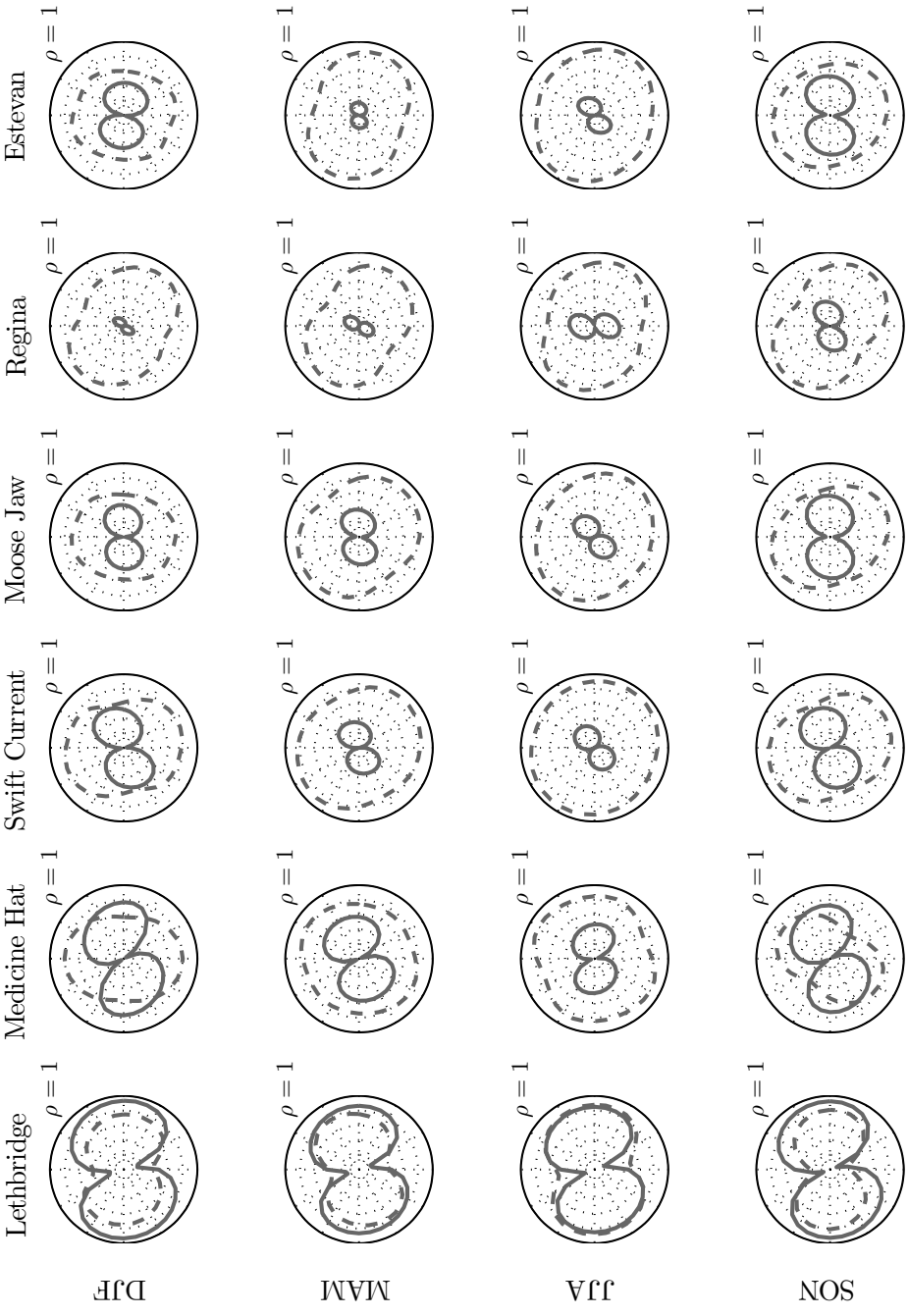

Figure C.30: As with Figure 4.11 for monthly-averaged wind quantities (across southern prairies sites).

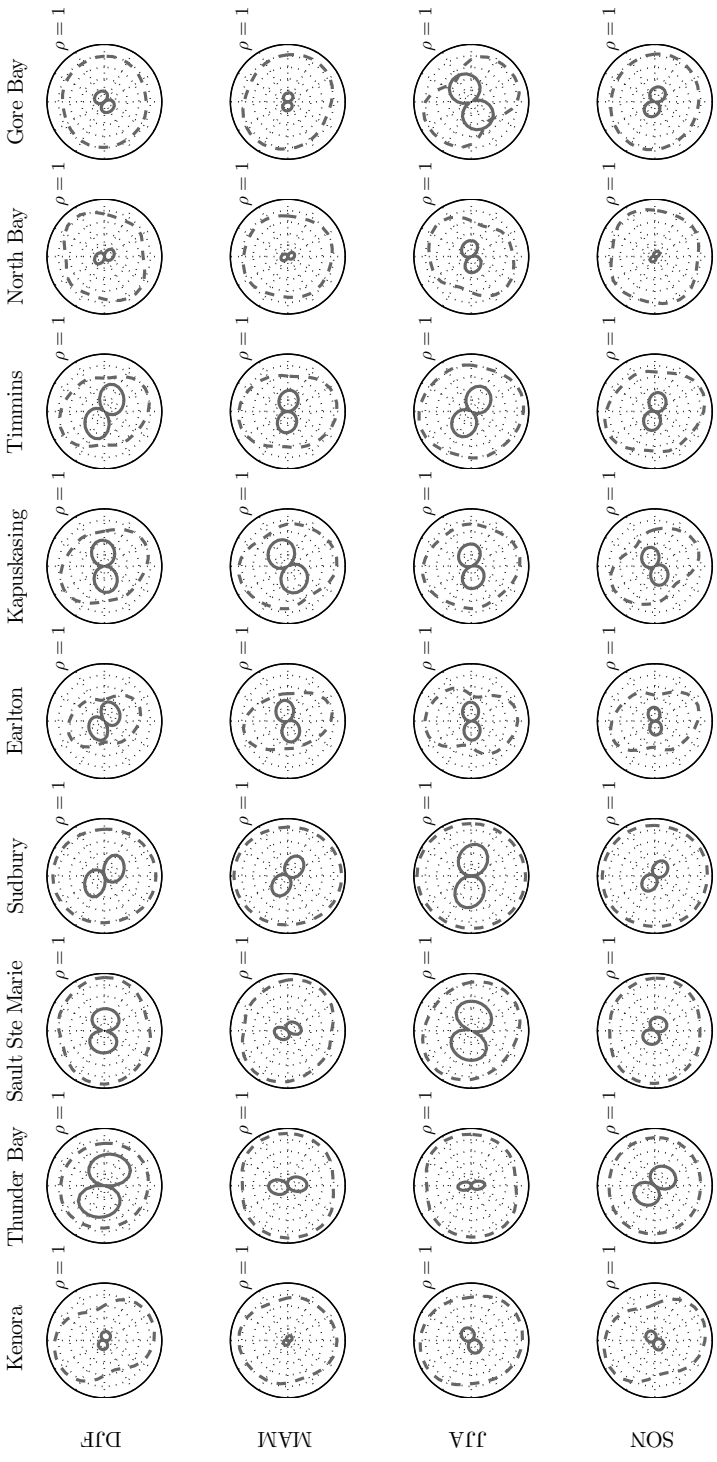

Figure C.31: As with Figure 4.11 for monthly-averaged wind quantities (across Ontario sites - 1 of 2).

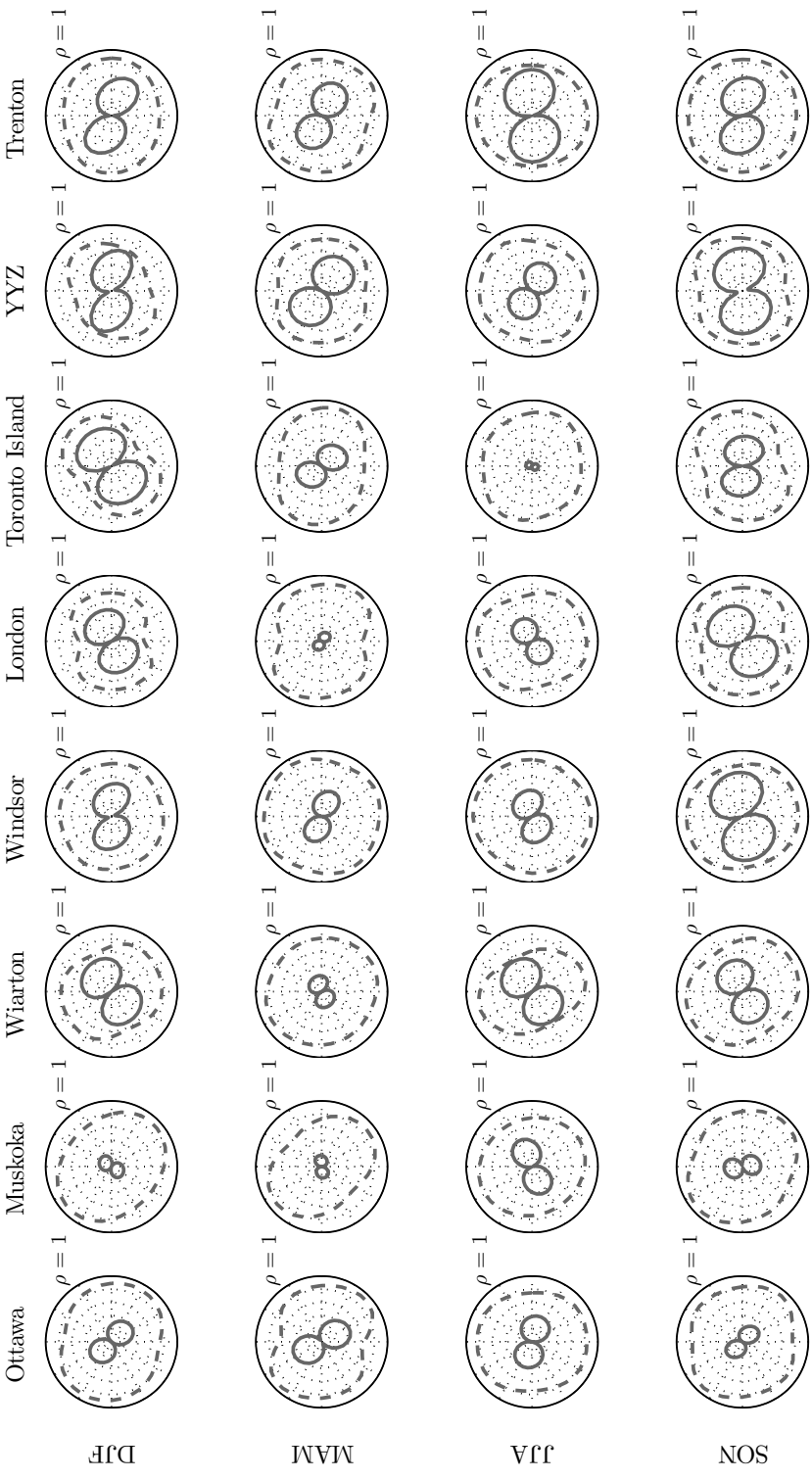

Figure C.32: As with Figure 4.11 for monthly-averaged wind quantities (across Ontario sites - 2 of 2).

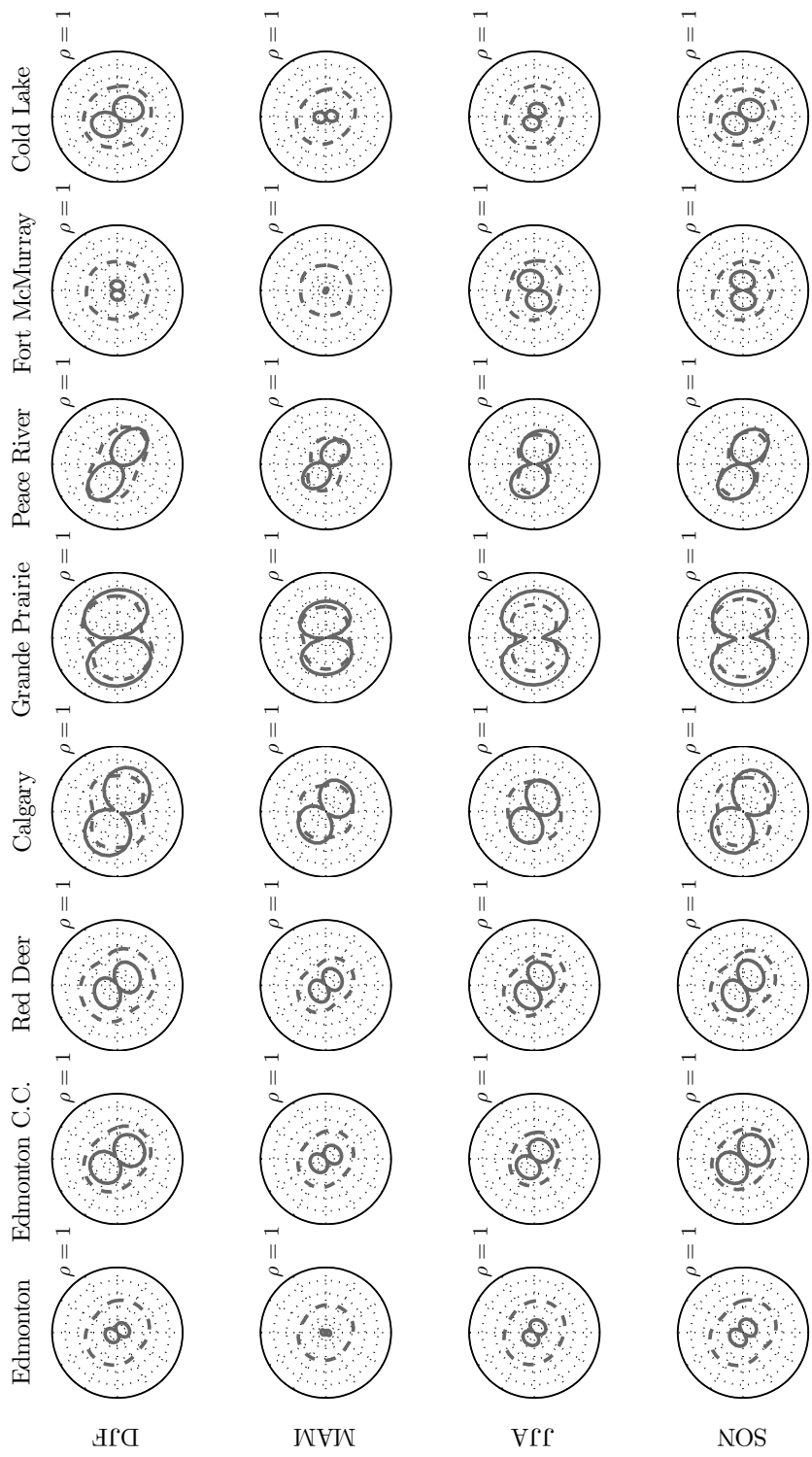

Figure C.33: As with Figure 4.11 for daily-averaged wind quantities (across northern plains sites).

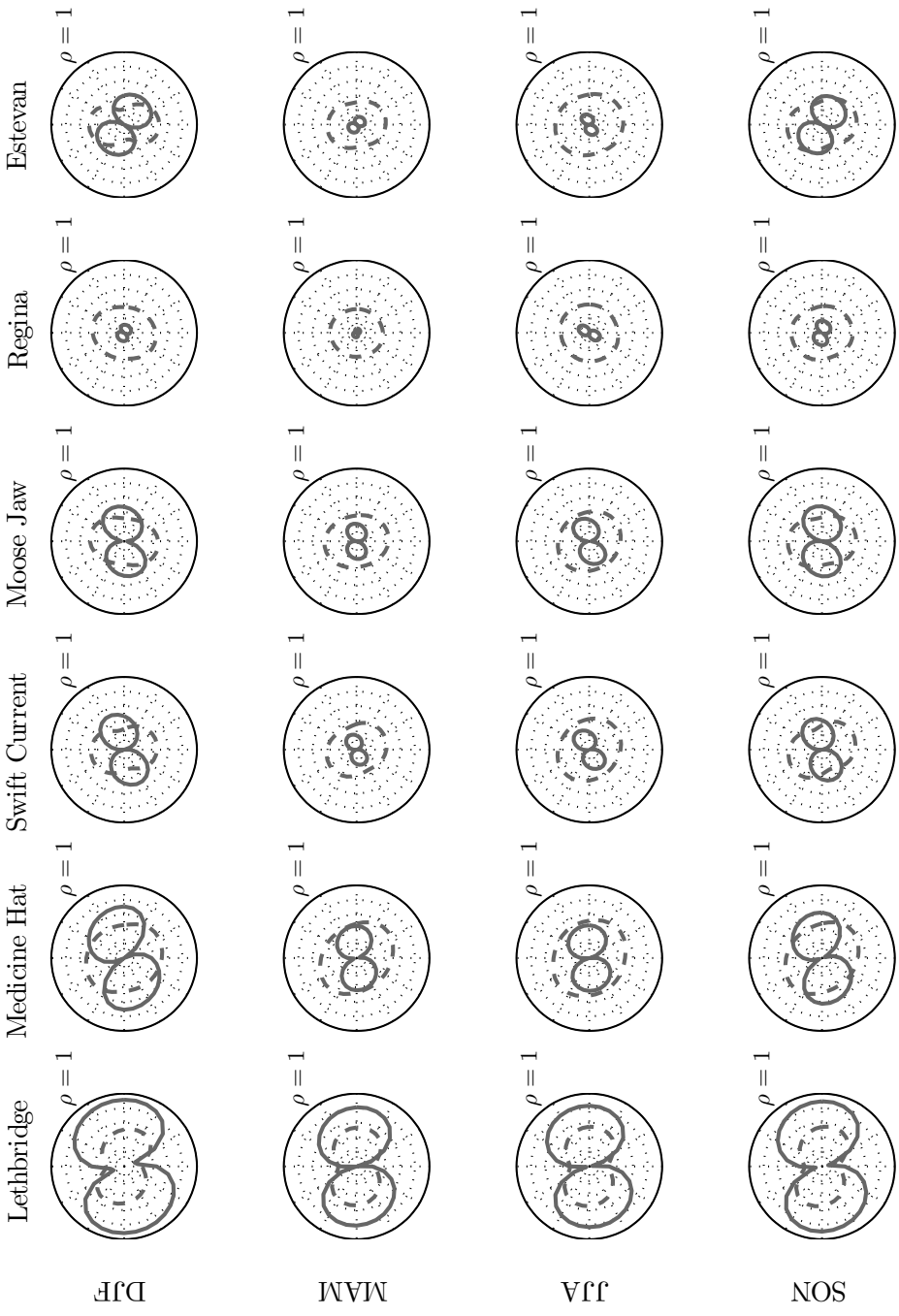

Figure C.34: As with Figure 4.11 for daily-averaged wind quantities (across southern prairies sites).

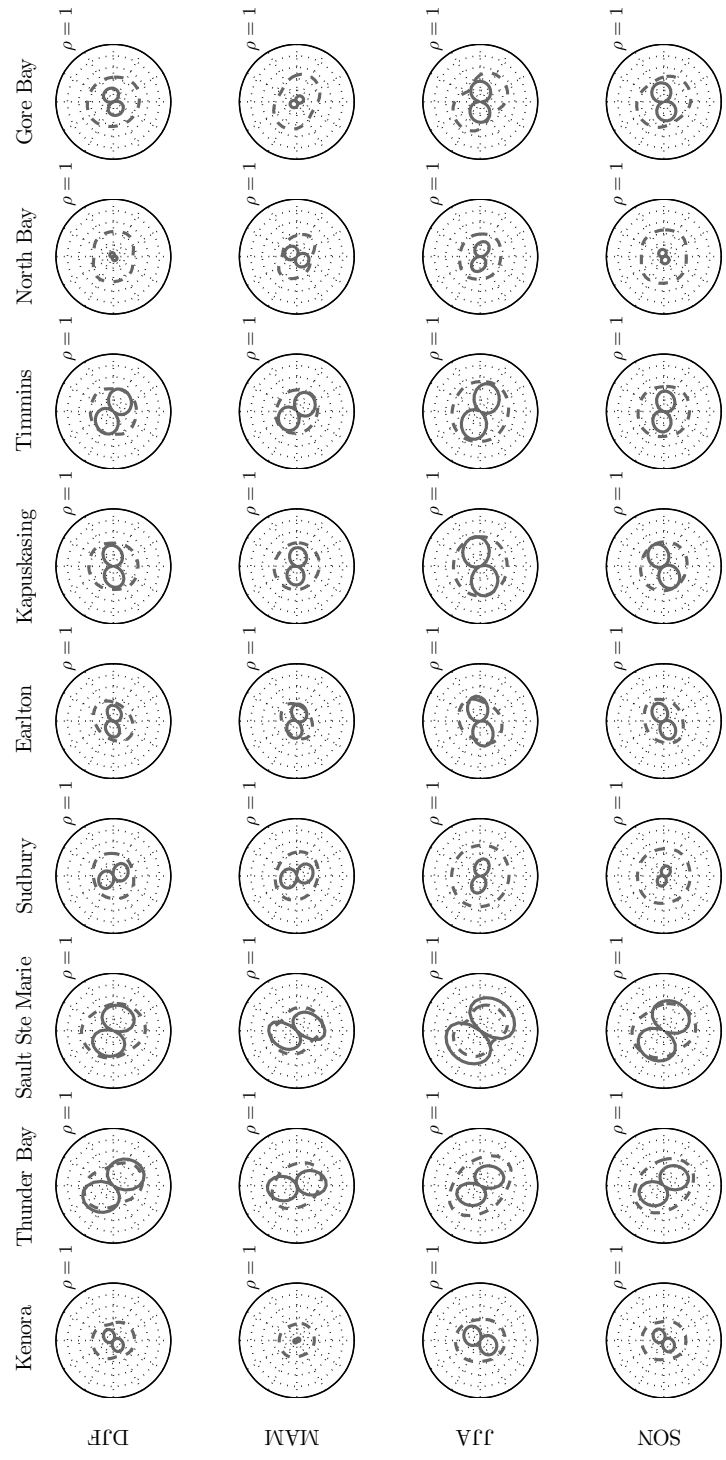

Figure C.35: As with Figure 4.11 for daily-averaged wind quantities (across Ontario sites - 1 of 2).

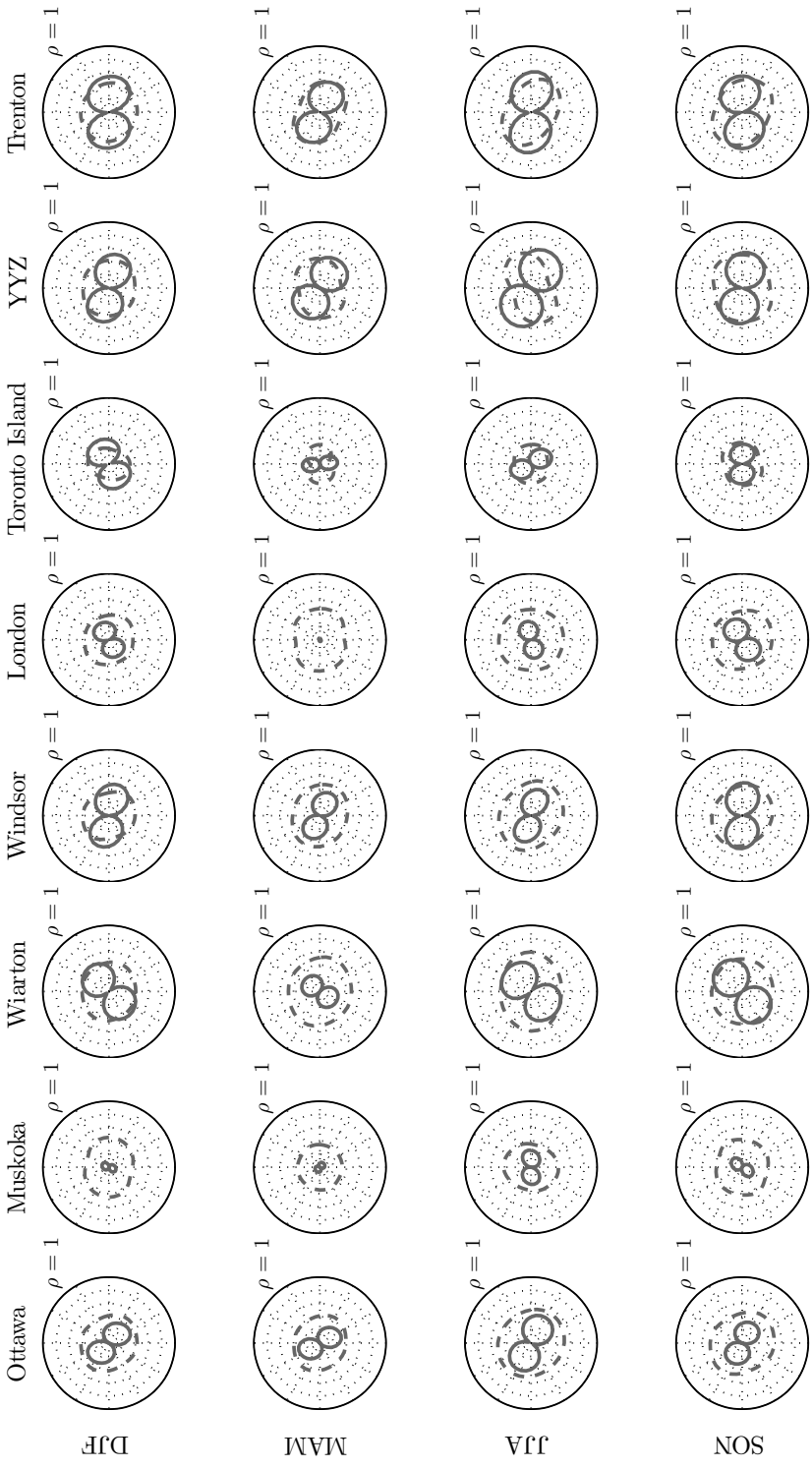

Figure C.36: As with Figure 4.11 for daily-averaged wind quantities (across Ontario sites - 2 of 2).