Ruprecht-Karls-Universität Heidelberg
Fakultät für Mathematik und Informatik

Ensemble Model Output Statistics for Wind Vectors

Diplomarbeit

von

Nina Schuhen

Betreuer: Prof. Dr. Tilmann Gneiting
Dr. Thordis L. Thorarinsdottir

September 2011

Zusammenfassung

Through the advent of ensemble forecasts, weather prediction experienced a systematic shift from the traditional deterministic view to a modern probabilistic approach. These ensembles, multiple runs of one or more mathematical models, however are often underdispersive, as only some of the uncertainties involved in numerical weather prediction are captured. Therefore statistical postprocessing methods were developed to transform discrete ensemble output into calibrated and sharp predictive distributions. So far these techniques only applied to one-dimensional weather quantities like temperature, air pressure or precipitation. In this thesis, we present a bivariate approach to address ensemble forecasts of two-dimensional wind vectors, extending the well-established EMOS method. We preserve its essential benefits and model the correlation between the vector components as a trigonometric function of the predicted wind direction. Additionally, our new method is tested against other forecasting techniques, using the UWME ensemble over the North American Pacific Northwest.

Zusammenfassung

Durch das Aufkommen von Ensemblevorhersagen hat die Wettervorhersage einen systematischen Wandel weg vom traditionell deterministischen und hin zu einem modernen probabilistischen Ansatz erlebt. Diese Ensembles, mehrere Durchläufe eines oder mehrerer mathematischer Modelle, sind jedoch oft unterdispersiv, weil nur einige der Unsicherheiten, die mit numerischer Wettervorhersage einhergehen, berücksichtigt werden. Daher wurden statistische Aufbereitungsmethoden entwickelt, um den diskreten Ensembleoutput in kalibrierte und scharfe Wahrscheinlichkeitsverteilungen umzuwandeln. Bisher ließen sich diese Verfahren jedoch nur auf eindimensionale Wettergrößen wie Temperatur, Luftdruck oder Niederschlag anwenden. In dieser Arbeit stellen wir eine bivariate Herangehensweise vor, die die bekannte EMOS Methode erweitert und sich mit Ensemblevorhersagen für zweidimensionale Windvektoren beschäftigt. Dabei behalten wir die grundlegenden Eigenschaften des univariaten Verfahrens bei und modellieren die Korrelation zwischen den Vektorkomponenten als eine trigonometrische Funktion der vorhergesagten Windrichtung. Des Weiteren führen wir eine Fallstudie für das UWME Ensemble über dem nordamerikanischen Pazifischen Nordwesten durch und testen unsere neue Methode gegen andere Vorhersageverfahren.

Contents

Chapter 1

Introduction

Prediction is very difficult, especially about the future.

Often attributed to Nobel laureate Niels Bohr (1885-1962)

Although predicting the future might be difficult, modern forecasting is perceived as a matter of great importance. In the process of making statements and inferences about the things to come, we draw on experiences from the past, often in form of statistics. Over the last decades, major aims of statistical analysis have become not only to make forecasts, but also to measure the associated uncertainty. Therefore, they should be expressed as probability distributions (Dawid 1984).

Many forecasts, however, are of deterministic nature, may it be for reasons of communication or decision making (Gneiting 2011). Still, a transition to probabilistic forecasts is in progress and statisticians function as a driving force (Gneiting 2008). The goal is not to eliminate, but to assess and evaluate uncertainty.

In meteorology, the probability of precipitation has been an integral part of most weather forecasts for a number of years. Yet, people often don't understand how to interpret such a percentage (Gigerenzer et al. 2005). But especially in this field, the probabilistic approach is very effective for a number of practical applications. Probabilities of exceeding a certain

threshold are used for issuing severe weather warnings, maintaining air and shipping traffic regulations, or for efficient agriculture as the risk of freezing temperatures. Palmer (2002) points out that weather and climate predictions have an enormous potential economic value: Financial profit can be traded off against the probability of heavy losses due to e.g. extreme weather.

Customers often have the choice between multiple forecasting products. But how do we rank rivalling forecasters? How can the "best" prediction be determined? Gneiting et al. (2007) proposed the principle of maximising the sharpness of a probabilistic forecast subject to its calibration. Calibration refers to the reliability of the forecast, i.e. the statistical consistency between the predictive probability distribution and the actually occurring observations. Sharpness quantifies the concentration of the distribution; under the condition that all forecasts are calibrated, we define the sharpest forecast to be the best.

Numerical weather prediction (NWP) simplified the process of incorporating probabilistic methods into traditional weather forecasting. The grid-based NWP models consist of a number of dynamical partial differential equations which are discretised and integrated forward to obtain future states of the atmosphere. The current state is assimilated from actual observations and provides the initial and lateral boundary conditions for these models.

Until the 1990s, one mathematical model was supplied with the one set of input data, which was deemed best to represent the current atmosphere (Gneiting and Raftery 2005). Chaos theory, however, implies that little deviations in the initial conditions can lead to significant forecast errors, as the observations are subject to a certain degree of imprecision. One reason for this is that observational locations are irregularly placed, and there are many areas with only sparse observational data, especially over the oceans. Interpolation to the grid structure therefore causes inherent uncertainties, which again limits the predictability, however skillful the model might be (Grimit 2001).

NWP models, on the other hand, became more and more accurate over time. In order to capture small-scale processes, the grid resolution was increased constantly. For instance, the COSMO-DE model of the German Weather Service (DWD), operational since 2007, has a horizontal spacing of 2.8 km (Baldauf et al. 2011). Despite the latest developments, however, there are limitations to the computer resources which are needed for such models.

Today, numerical weather prediction has changed from being a deterministic matter to the probabilistic approach of ensemble forecasting (see e.g. Leutbecher and Palmer 2008). Ensembles consist of multiple point forecasts, named ensemble members, generated by NWP models and can be ordered into three groups:

- *Multianalysis ensembles:*
  One numerical model is run multiple times, using slightly different sets of initial conditions.

- *Multimodel ensembles:*
  Multiple numerical models are run with a single set of initial conditions.

- *Multimodel multianalysis ensembles:*
  As a combination of the previous ones, multiple sets of initial conditions are used to run multiple numerical models.

In this way, the two major uncertainty sources are addressed: the imperfections in the initial conditions and the inadequacies of the mathematical models. Ensemble prediction system have been implemented for both short-range (0-48 hours) and medium-range (2-10 days) weather forecasting (Eckel and Mass 2005), in the form of regional and global NWP models, respectively. Figure 1.1 shows a schematic representation of an exemplary multimodel multianalysis ensemble with two models and eight ensemble members. The ensemble pictured in Figure 1.2 is the multianalysis University of Washington Mesoscale Ensemble (UWME), as operational in 2008. A full description can be found in Section 5.1.

The context of ensemble forecasts provides the possibility of combining a point forecast, say the ensemble mean, with an estimation of its accuracy, through for instance the root mean squared error. On average, the ensemble mean outperforms each of the individual ensemble members, in that it is the best estimate for the verifying state of the atmosphere (Grimit and Mass 2002).

Another important aspect is the so called spread-error correlation or spread-skill relationship of the ensemble. Often there exists a link between the a priori known ensemble spread (or ensemble variance) and the observed error of the ensemble mean forecast. Whitaker and

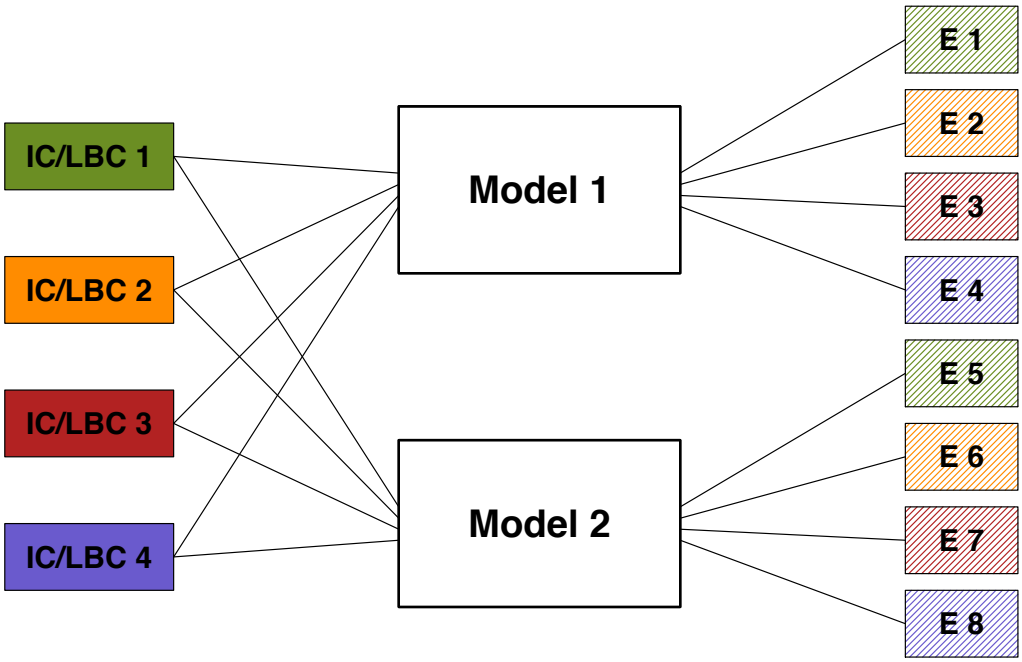


Figure 1.1: Diagram of an exemplary multimodel multianalysis ensemble forecast; every set of initial and lateral boundary conditions IC/LBC 1 - IC/LBC 4 is integrated forward through the two numerical models, resulting in eight ensemble members E 1 - E 8

Loughe (1998) and Grimit (2001) noticed that especially extremely large or small spread is a good estimator for the eventual forecast skill.

Despite this desirable property, ensemble predictions were found to be uncalibrated (see e.g. Hamill and Colucci 1997), in that they are underdispersive. This means that the ensemble spread is on average too small and the observation lies too often outside of the ensemble range. Raftery et al. (2005) reason that ensembles only capture some of the uncertainties involved in numerical weather forecasting, and those only partially. Moreover, many weather variables are continuous quantities and should be expressed in terms of continuous probability distributions instead of discrete ensembles of a finite size (Gneiting et al. 2005).

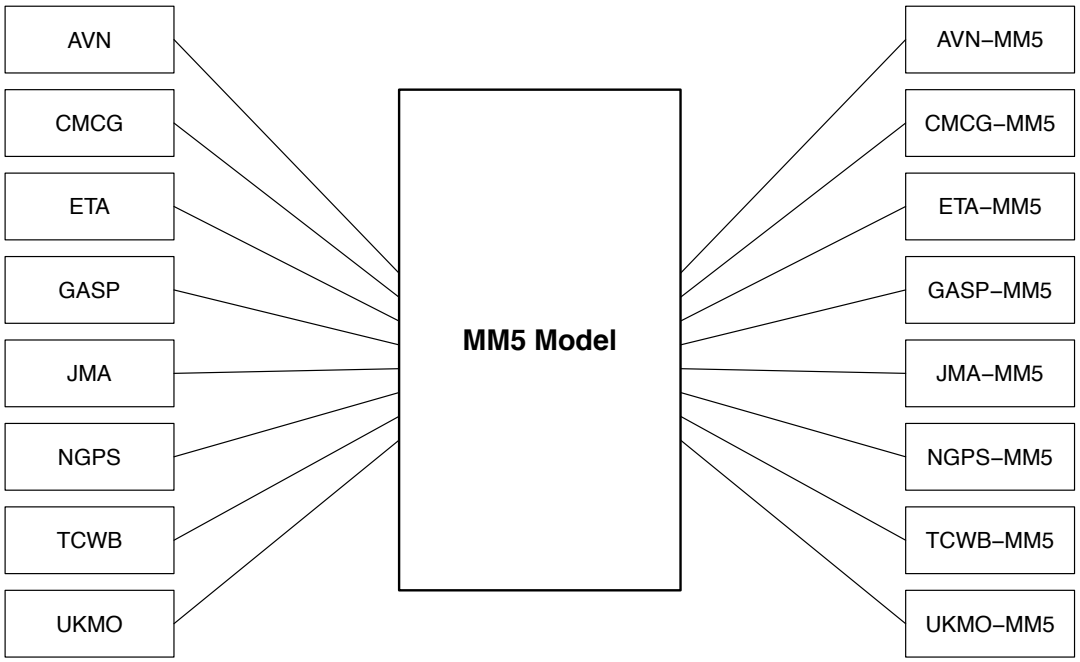


Figure 1.2: Diagram of the UWME; eight different sets of initial and boundary conditions (on the left) are integrated forward using the MM5 numerical model, resulting in eight ensemble members (on the right)

All these problems, as well as model biases and different spatial resolutions of forecast grid and observations, can be addressed with statistical postprocessing (Gneiting and Raftery 2005). Several such methods have been suggested, with ensemble model output statistics (EMOS) and Bayesian model averaging (BMA) being the most prominent. EMOS, elaborately discussed in the next chapter, is based on multiple linear regression and fits a normal distribution to the ensemble member forecasts. BMA assigns probability density functions to the individual ensemble members and generates a weighted average of these densities, where the weights reflect the forecasting skill of the respective ensemble members. More information about BMA can be found in Raftery et al. (2005), Sloughter et al. (2007), Sloughter et al. (2010) and Bao et al. (2010).

Both methods apply to a number of different weather variables like temperature, air pressure or wind speed, however these are only univariate quantities. Surface wind, in contrast, is measured in two dimensions, either as zonal and meridional vector components or as wind speed and wind direction. The application of wind vector forecasts ranges over many different fields: Air traffic control relies on short-ranged local and regional forecasts, not only to warn of extreme weather, but also to optimise the coordination of runways. For ship routeing, especially probabilistic wind forecasts have a huge economic value, as it can avoid delays and save fuel. Finally, one of the most important purposes is the production of wind energy; here, the combination of wind speed forecasts, to estimate the production amount, and wind direction forecasts, to adjust the turbines, plays a significant role.

So far, those two variables were always addressed separately (Bao et al. 2010; Sloughter et al. 2010; Thorarinsdottir and Gneiting 2010), not taking into account a possible relationship. For the above mentioned applications, particularly the correlation structure might be of crucial consequence. Therefore we propose a new approach to postprocess ensemble forecasts of wind vectors jointly, so that spatial properties are inherited from the underlying numerical models. We base our method on the aforementioned EMOS and extend it to bivariate weather quantities, where the correlation between the two vector components is estimated from historic data.

Current research relating to wind vectors includes a two-dimensional variant of BMA (Sloughter 2009) and a technique to adaptively calibrate wind vector ensembles (Pinson 2011). The Ensemble Copula Coupling approach by Schefzik (2011) refers to multidimensional forecasts in general and can easily be adapted to our task of postprocessing wind vectors.

The new EMOS method will be described and tested as follows:

In Chapter 2, we present several methods to assess the predictive skill of both univariate and multivariate probabilistic forecasts, whether in the form of ensembles or probability distributions. These tools will not only be used to compare performances of forecasts, but are also a part of the parameter estimation, and make sure that the resulting probabilistic forecasts are calibrated and sharp. Here, we refer to Gneiting et al. (2008), where these multidimensional assessment tools were described.

In Chapter 3, the conventional EMOS postprocessing method is discussed in detail. We show all its adaptations and applications and point out the special characteristics and key elements, which will later be incorporated into the bivariate approach.
In Chapter 4, we develop the new bivariate method, beginning with an empirical analysis of data provided by the UWME ensemble (see Section 5.1). We then propose a model for the bivariate predictive distribution and split the parameter estimation into three parts, thereby predicting forecast mean, correlation and spread successively.
In Chapter 5, several other forecasting and postprocessing techniques suitable for wind vectors are reviewed and we compare the results in form of a case study. The performance of these methods is assessed for the North American Pacific Northwest in 2008, using the tools from Chapter 2. For implementing the methods described in this thesis and in order to create the graphics, the `R` environment for statistical computing is employed. More information can be found in R Development Core Team (2011).
Finally, in Chapter 6, we provide a summary of the new EMOS method for bivariate wind vector forecasts and discuss the results of the case study. Also, future prospects are shown, especially in terms of wind field forecasting.

Chapter 2

Assessing Forecast Skill

In this chapter, we will present several tools that can be used to assess the forecast skill of the new EMOS method and of the rivalling forecasting techniques described in Section 5.3. In Gneiting et al. (2008), it was established how multivariate ensemble or density forecasts can be evaluated, so this article will function as a guideline for choosing which tools to use. Often there exists a multidimensional extension to well-known one-dimensional methods, and we will discuss both univariate and multivariate assessment.

Due to the fact that the thesis at hand concerns wind vectors, we will here focus on forecast assessment in two dimensions. Furthermore, as the new EMOS method will also be used to produce wind speed forecasts (Section 5.5), the univariate tools are employed in the case study as well.

Additional to the purpose of assessing forecast skill, so called proper scoring rules are a key element of the EMOS postprocessing method itself. As mentioned in the preceding chapter, the goal of probabilistic forecasting is to maximise the sharpness subject to calibration (Gneiting et al. 2007). Proper scoring rules, elaborately discussed in Gneiting and Raftery (2007), provide a way to address both of these properties simultaneously, so a natural conclusion would be to involve them in the parameter estimation. Minimum CRPS estimation, for instance, is used for the so far known univariate EMOS methods, and we want to find a similar approach for the bivariate method.

Probabilistic forecasts occur in the form of discrete ensembles or probability densities and there are tools that apply to one or the other. However, it does not pose difficulties to

draw samples from a predictive distribution and thus create a forecast ensemble, or to fit an appropriate distribution to given ensemble forecasts. In this manner, both types of forecasts can easily be transformed into one another. More details are to be found in Section 5.3.5.

First we take a look at tools to check the calibration of probabilistic forecasts, then the quantification of sharpness will be addressed. In the final section we will introduce several proper scoring rules, which again will later be used to determine the parameters of the EMOS predictive distribution.

2.1 Assessing Calibration

Calibration, sometimes referred to as reliability, is a measure for the statistical consistency between the probabilistic forecast and the verifying observation. Simply said, this means that if some event is predicted to occur with a certain probability, say 40%, on average it should happen in about 40% of all times. Thus it is not only a property of the probabilistic forecasts, but also of the verifying observations.

In order to assess the calibration of a forecast, a probability integral transform (PIT) histogram is usually employed for predictive distributions, while its ensemble forecast counterpart is called verification rank histogram (VRH).

2.1.1 Univariate Histograms

The preferred tool for univariate forecast distributions, the *PIT histogram* (Dawid 1984; Diebold et al. 1998), is based on the assumption that an observed value x can be perceived as randomly sampled from a "true" distribution G (Gneiting et al. 2007). If the predicted distribution F is indeed identical to this true distribution, the value of the predictive cumulative distribution function (CDF) at x has a uniform distribution

$$p = F(x) \sim \mathcal{U}[0, 1].$$

Therefore, for all forecasts available, the PIT values p are computed, sorted into bins and then formed into a histogram. In this histogram, deviations from uniformity are easy to

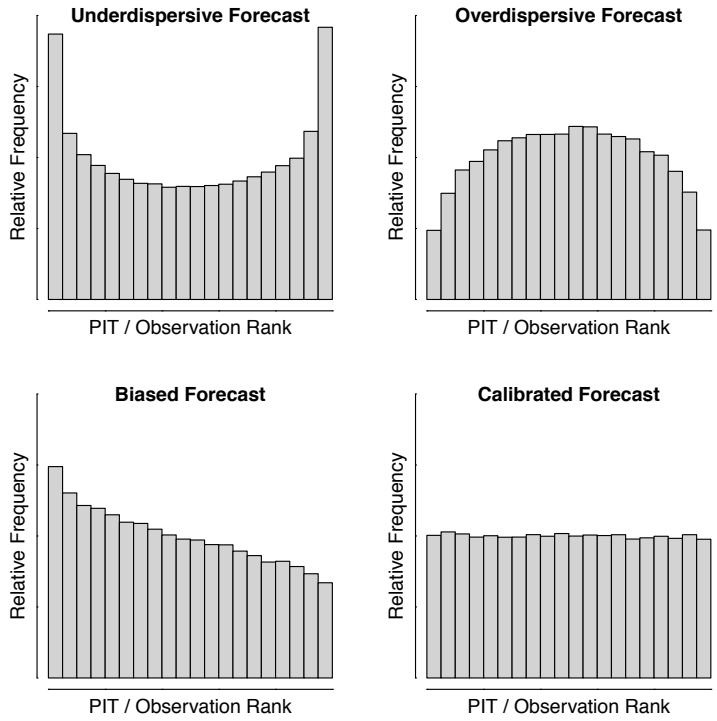


Figure 2.1: Examples for different shapes of the PIT or verification rank histogram

recognise and one can interpret these deviations. An U-shaped histogram indicates that the spread of the underlying predictive distribution is too small, it is underdispersed. On the other side, an overdispersed distribution reveals itself in a hump-shaped diagram, with too many observations in the center of the distribution. Biases can result in a skewed or triangle-shaped histogram, while a nearly flat histogram suggests calibration.

Hamill (2001) however proved that the latter case can be misleading: Although a uniform PIT (or rank histogram, for that matter) is required for a probabilistic forecast to be calibrated, it is not sufficient. To address this issue, Gneiting et al. (2007) proposed the principle of maximising the sharpness subject to calibration, so that the performance of several forecasts can be evaluated by comparing the respective sharpness measures, under the condition of them being calibrated.

In the context of a discrete ensemble forecast, the *verification rank histogram* or *Talagrand diagram* (Anderson 1996; Hamill and Colucci 1997; Talagrand et al. 1997) takes the role of the PIT histogram. It relies on the assumption that for a calibrated ensemble of size M, the observation has the same probability to occupy any rank between 1 and $M+1$ when pooled with the ordered ensemble members. This means, for each time and location, we order the ensemble forecasts and observation, find the rank of the observation and finally plot the histogram of the aggregated ranks. As before, a U-like shape implies underdispersion, while the opposite holds for a hump-like shape. Skewed histograms indicate a certain bias and calibrated ensembles produce a flat histogram. In Figure 2.1, four exemplary histograms are shown.

For comparing different forecasting techniques, it can be useful to quantify the calibration. This is possible with the so-called *discrepancy* or *reliability index* Δ (Delle Monache et al. 2006; Berrocal et al. 2007), which is gathered from the respective histograms,

$$\Delta = \sum_{i=1}^{M+1} \left| f_i - \frac{1}{M+1} \right|.$$

It measures the deviation between the observed relative frequency f_i for bin i and the desired value $\frac{1}{M+1}$, accumulated over all bins.

2.1.2 Multivariate Rank Histogram

A natural generalisation of the VRH to multiple dimensions is the *multivariate rank histogram* (MRH), the only challenge lies in defining a multivariate rank order. For simplicity reasons, the description below is given for two dimensions, it is however easily applicable to ensemble forecasts of any dimension.

First, we write for two vectors $(x_1, x_2)^T$ and $(y_1, y_2)^T$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \preceq \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \text{ if and only if } x_1 \leq y_1 \text{ and } x_2 \leq y_2.$$

If we consider the ensemble forecast $\{\mathbf{x}_j \in \mathbb{R}^2 : j = 1, \ldots, M\}$ and its respective verifying observation $\mathbf{x}_0 \in \mathbb{R}^2$, we proceed according to the following steps:

1. *Standardise*:
   It is often useful to apply a principal component transformation to the set $\{\mathbf{x}_j : j = 0, \ldots, M\}$, which contains forecasts and observation, thus resulting in standardised values $\left\{\mathbf{x}_j^\star : j = 0, \ldots, M\right\}$. In the current context, we refrain from doing so.

2. *Assign pre-ranks:*
   In the next step, we determine pre-ranks ρ_j for the possibly standardised values $\mathbf{x}_j^\star$ and all $j = 0, \ldots, M$. The pre-rank is defined as

   $$\rho_j = \sum_{k=0}^{M} \mathbb{I}\left\{\mathbf{x}_k^\star \preceq \mathbf{x}_j^\star\right\},$$

   where $\mathbb{I}$ denotes the indicator function. In this way, we find, for each vector from the combined set of ensemble member forecasts and observation, the number of vectors from the same set which are smaller or equal in each vector component. The pre-ranks therefore are integers between 1 and $M + 1$.

3. *Obtain the multivariate rank of the observation:*
   For the multivariate rank r, we note the rank of the observation pre-rank, while possible ties are resolved at random. We write

   $$s^< = \sum_{j=0}^{M} \mathbb{I}\{\rho_j < \rho_0\} \quad \text{and} \quad s^= = \sum_{j=0}^{M} \mathbb{I}\{\rho_j = \rho_0\}$$

   as the number of pre-ranks that are smaller than the pre-rank of the observation and the number of those that are equal to it. Then r is chosen randomly as any integer between $s^< + 1$ and $s^< + s^=$, or more specifically, from a discrete uniform distribution on $\{s^< + 1, \ldots, s^< + s^=\}$. Again, r can range from 1 to $M + 1$.

4. *Aggregate ranks and plot histogram:*
   For the final step, we aggregate the multivariate ranks over all forecast dates and locations available and plot the histogram of these ranks. To reduce random

variability caused by sampling the multivariate rank, the histogram intensities are averaged over 100 runs.

Figure 2.2 illustrates the process of determining first the pre-ranks and then the multivariate observation rank. Four ensemble member forecasts lie to the lower left of the observation vector, i.e. are smaller in each of the vector components. The observed value therefore has the pre-rank 5, the same as two other ensemble members. The pre-ranks of five vectors are smaller than that of the observation, so we sample the multivariate rank from the set $\{s^{<}+1,\ldots,s^{<}+s^{=}\}=\{6,7,8\}$. In this case, r was chosen to be 6.

The multivariate rank histogram can be interpreted in the same manner as its univariate counterpart. We will use this tool for assessing both ensemble and density forecasts, where predictive PDFs will be converted to discrete ensembles as described in Section 5.3.5. The reliability index Δ provides a measure for the deviation from uniformity, in the same manner as for the univariate histograms.

2.1.3 Marginal Calibration Diagram

Gneiting et al. (2007) refer to marginal calibration as the equality of forecast climate and observed climate. If the forecast and the observation can be assumed compatible, their respective empirical distributions, aggregated over a number of cases in a steady state, should be statistically consistent.

In the context of the thesis at hand, we check this property by a location-specific diagram. For each observational location, the empirical distributions of the raw ensemble, the EMOS method and the observations are plotted over the course of one calendar year. To represent the respective empirical distributions of the forecasts, for each date one random ensemble member and one random vector from the EMOS predictive distribution are drawn.

We then say that a forecast is marginally calibrated if the pattern created by the wind vector observations is approximately reproduced by the distribution of the forecast. The use of this so-called *marginal calibration diagram* can reveal forecast biases or dispersion errors and especially address the possible influence of local terrain and surroundings.

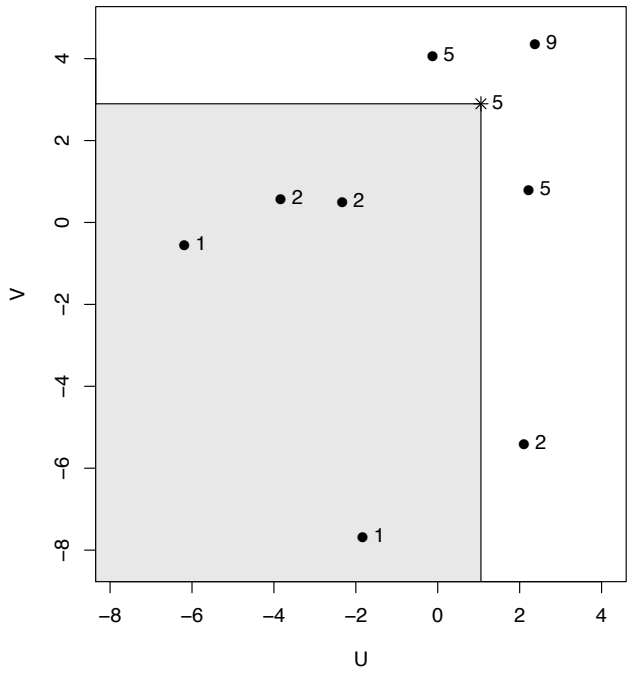


Figure 2.2: Example of how to determine the pre-ranks and the multivariate rank for the MRH; pictured is an ensemble forecast for wind vectors, created by sampling 8 vectors from the EMOS predictive distribution for Sea-Tac airport valid on 23 January 2008, and the respective verifying observation, indicated by the star; for each vector, the pre-rank is given as the number next to it; the forecast itself is described in chapters 4 and 5

2.2 Assessing Sharpness

After confirming that they are calibrated, it is necessary to measure the sharpness of competing forecasts so that their quality can be compared. While calibration provides information about the reliability of a probabilistic forecast, sharpness refers to its concentration and is therefore a property of the forecast only. The sharper a calibrated predictive distribution, the less the amount of uncertainty and the better the performance.

For univariate quantities, the sharpness measure of our choice is the standard deviation of both predictive PDFs and ensemble forecasts. In the two-dimensional case, we employ a generalisation of the standard deviation, the *determinant sharpness* (DS)

$$\begin{aligned} \mathrm{DS} &= (\det \boldsymbol{\Sigma})^{1/4} \\ &= \left(\sigma_1^2 \sigma_2^2 \cdot \left(1 - \rho^2\right)\right)^{1/4}, \end{aligned}$$

with $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\,\sigma_1\sigma_2 \\ \rho\,\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ being the variance-covariance matrix of the forecast. This tool is also applicable to discrete ensembles, where the matrix is generated by the empirical variances and correlation of the ensemble.

However, the determinant sharpness has a major disadvantage, which was brought up by Jolliffe (2008). For nearly singular matrices, such as in the case of highly correlated variables, it becomes very small, although the individual variances may be large. Therefore it is possible for such forecasts to seem sharper than those with smaller spread and correlation. This problem indeed arises in our case study and is discussed in Section 5.4.3.2.

2.3 Proper Scoring Rules

Additional to measuring the sharpness and calibration, the performance of a forecast can be evaluated by assigning a numerical score. For this we introduce a function $s(P, \mathbf{x})$, where P is the predictive distribution and $\mathbf{x}$ the observed vector or value. This function, called *scoring rule*, is negatively oriented and can be taken as a penalty. If we assume $\mathbf{x}$ to be drawn from a distribution Q, the expected value is written as $s(P, Q)$. The idea of proper scoring rules and the mathematical background are discussed in Gneiting and Raftery (2007).

Naturally, the intention of a forecaster should be to minimise the penalty by issuing the best forecast he is able to make (Savage 1971; Bröcker and Smith 2007). Thus the score becomes the smallest when the true distribution is predicted. We say that a scoring rule is proper if

$$s(Q, Q) \leq s(P, Q)$$

holds for all P and Q, and strictly proper with equality if and only if $P = Q$. This property is very important for the assessment of our forecasts, meaning that a proper scoring rule addresses calibration and sharpness simultaneously by giving a measure for the overall skill (Winkler 1977, 1996).

In practical applications, the reported values are typically averaged over a period of time or for a specific location. We note these mean scores as

$$S_n = \frac{1}{n} \sum_{i=1}^{n} s\left(P_i, \mathbf{x}_i\right),$$

where P_i and $\mathbf{x}_i$ range over all forecast cases $i = 1, \ldots, n$.

2.3.1 Univariate Forecasts

One of the scoring rules coming into use in the new EMOS method is the widely known *logarithmic* or *ignorance score* (Good 1952; Bernardo 1979). It can be applied to univariate as well as multivariate density forecasts. For a predictive distribution P with density function p and the observed vector or value $\mathbf{x}$, the ignorance score is defined as

$$\operatorname{logs}\left(P, \mathbf{x}\right) = -\log p\left(\mathbf{x}\right).$$

However, this proper scoring rule can assign infinite penalties, which poses problems in practice. The so-called *continuous ranked probability score* (CRPS), proposed in Matheson and Winkler (1976), is more robust and also more versatile as it applies to density and ensemble forecasts. Gneiting and Raftery (2007) showed that the following two representations of the CRPS are equal, conditional on P having finite first moment:

$$\operatorname{crps}\left(P, x\right) = \int_{-\infty}^{\infty} \left(F\left(x\right) - \mathbb{I}\left\{y \geq x\right\}\right)^2 \mathrm{d}y \tag{2.1}$$

$$= \mathbb{E}_P \left|X - x\right| - \frac{1}{2} \mathbb{E}_P \left|X - X'\right|. \tag{2.2}$$

Here, F denotes the cumulative distribution function of P, while X and X' are independent random variables with distribution P. In contrast to the ignorance score, the observation x may only be a number rather than a multidimensional vector. The CRPS is reported in the same unit as the observation which simplifies interpretability.

If we consider deterministic instead of probabilistic forecasts, we are interested in the predictive skill of an ensemble or a predictive distribution when issuing a point forecast μ. According to Gneiting (2011), the most effective way of producing point forecasts is to specify a scoring rule in advance and then choose the optimal point predictor for the scoring rule at hand, i.e. the Bayes rule. In the situation above, P can be taken as a point measure δ_μ and the CRPS reduces to the *absolute error*

$$\mathrm{ae}\,(P, x) = |\mu - x|\,.$$

The respective Bayes rule for the absolute error is the median of the predictive distribution, med_P. We will use the mean absolute error (MAE) of the forecast median and the mean CRPS mainly for testing the skill of the new EMOS method for wind vectors when transformed into wind speed forecasts (Section 5.5).

2.3.2 Multivariate Forecasts

The direct generalisation of the CRPS to multiple dimensions is called the *energy score* (ES), introduced by Gneiting and Raftery (2007). It derives from the kernel score representation (2.2) and uses the Euclidean norm $\|\cdot\|$ instead of the absolute value:

$$\mathrm{es}\,(P, \mathbf{x}) = \mathbb{E}_P\,\|\mathbf{X} - \mathbf{x}\| - \frac{1}{2}\,\mathbb{E}_P\,\|\mathbf{X} - \mathbf{X}'\|\,,$$

where $\mathbf{X}$ and $\mathbf{X}'$ are independent random vectors distributed with P and $\mathbf{x}$ is the observation vector.

For ensemble forecasts, where the predictive distribution P_{ens} places a point mass of $\frac{1}{M}$ on the ensemble members $\mathbf{x}_1, \ldots, \mathbf{x}_M \in \mathbb{R}^2$, the energy score can easily be evaluated as

$$\operatorname{es}(P_{\text{ens}}, \mathbf{x}) = \frac{1}{M}\sum_{j=1}^{M}\|\mathbf{x}_j - \mathbf{x}\| - \frac{1}{2M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\|\mathbf{x}_i - \mathbf{x}_j\|.$$

However, for most densities this proper scoring rule is not straightforward to compute. As we use bivariate normal distributions, a Monte Carlo approximation can replace the usual form of the energy score, reducing the computational effort notably. The energy score then becomes

$$\widehat{\operatorname{es}}(P, \mathbf{x}) = \frac{1}{k}\sum_{i=1}^{k}\|\mathbf{x}_i - \mathbf{x}\| - \frac{1}{2(k-1)}\sum_{i=1}^{k-1}\|\mathbf{x}_i - \mathbf{x}_{i+1}\|,$$

where $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is a simple random sample of size $k = 10,000$, drawn from the predictive distribution P.

Again, deterministic forecasts can also be assessed by this scoring rule. As in the case of the CRPS, we set P to be the point measure $\delta_{\boldsymbol{\mu}}$. The resulting score is called the *Euclidean error* (EE) and is the generalisation of the absolute error:

$$\operatorname{ee}(P, \mathbf{x}) = \|\boldsymbol{\mu} - \mathbf{x}\|.$$

As point forecast $\boldsymbol{\mu}$, we choose the appropriate Bayes rule, the spatial median med_P. Considering an ensemble forecast, it is defined as the vector that minimises the sum of the Euclidean distances to the ensemble members,

$$\sum_{i=1}^{M}\|\operatorname{med}_P - \mathbf{x}_i\|.$$

This vector can only be determined numerically, e.g. using the algorithm described in Vardi and Zhang (2000) and implemented in the `R` package `ICSNP`.

In this chapter different tools for the assessment of both ensemble and density forecasts were introduced. These tools will be used in Chapter 5 to analyse and compare the performance of several competing forecasting and postprocessing techniques. The univariate and the multivariate rank histograms and the PIT histogram help us to test forecasts for calibration, while the marginal calibration diagram unmasks local biases and dispersion errors. As to the

sharpness, the determinant sharpness provides a way to quantify the spread of multivariate distributions. Proper scoring rules are able to assess calibration and sharpness simultaneously and therefore can be employed for the parameter estimation. The conventional EMOS method for univariate weather quantities, for instance, described in the following chapter, relies on minimum CRPS estimation.

Chapter 3

EMOS for Univariate Weather Quantities

In Chapter 1, we already mentioned the well-established statistical postprocessing method EMOS, also called non-homogeneous Gaussian regression, and stated that such techniques can improve the quality and accuracy of weather forecasts significantly. EMOS was first proposed by Gneiting et al. (2005) and then developed further in Thorarinsdottir and Gneiting (2010) and Thorarinsdottir and Johnson (2011). Our goal is to find a two-dimensional extension which applies to wind vectors by capturing the essential properties of the one-dimensional variant. The idea and concept behind EMOS is the subject of this chapter.

3.1 General Idea

Standard model output statistics (MOS) techniques profit from determining a statistical relationship between the predictive weather quantity and the output variables of a numerical model (Glahn and Lowry 1972; Wilks 1995). For instance, the prediction of wind speed can correlate strongly with the wind vector forecast or with past wind speed observations. These are then combined as predictors in a multiple linear regression equation.

EMOS is a form of MOS or multiple linear regression; here, the ensemble members function as predictors. By fitting a normal distribution with the regression estimate as predictive

mean, and the mean squared prediction error of the training data as predictive variance, a full probability distribution can be gained from a regression equation. This straightforward way, however, does not take account of a potential relationship between ensemble spread and forecast skill (Whitaker and Loughe 1998). Therefore, the EMOS method models the predictive variance as an affine function of the ensemble variance.

This simple concept addresses forecast biases and dispersion errors on one hand and benefits from the skill of the underlying numerical models on the other. It therefore combines statistical and numerical weather prediction, while being very parsimonious and easy to implement.

3.2 Standard EMOS

In this section, we discuss EMOS as it was originally described by Gneiting et al. (2005). It applies to any weather quantity Y, whose distribution, conditional on the ensemble member forecasts $X_1, \ldots, X_M$, can be modelled with a normal distribution; such quantities are surface temperature, sea level pressure, or wind vector components (see Section 5.3.3).

For the predictive mean, the authors suggest a linear regression equation with the ensemble members as predictors and the observation Y as predictand,

$$Y = a + b_1 X_1 + \ldots + b_M X_M + \varepsilon.$$

The error term ε has distribution $\mathcal{N}(0, c + d\,S^2)$, where $S^2 = \frac{1}{M}\sum_{i=1}^{M}\left(X_i - \overline{X}\right)^2$ is the empirical variance of the ensemble and $\overline{X} = \frac{1}{M}\sum_{i=1}^{M} X_i$ the ensemble mean. The variance term corrects the underdispersion of the ensemble, while taking account of the spread-error correlation. Thus, the full predictive distribution for Y is denoted as

$$Y \mid X_1, \ldots, X_M \sim \mathcal{N}\left(a + b_1 X_1 + \ldots + b_M X_M, c + d\,S^2\right).$$

The coefficients $b_1, \ldots, b_M$ can be interpreted as the performance of the individual ensemble member models over the training period, relative to the other members. Additionally, highly correlated ensemble members can be identified, as typically the most skillful of them receives a substantial coefficient estimate, while the estimates of the others are nearly zero.

The spread parameters c and d reflect the relationship between the ensemble spread and the forecast error in the training data. If there is a significant correlation between these two, d tends to be larger and c smaller. On the other hand, when there is no information to be gained by the ensemble variance, d will be negligibly small and c very large.

For estimating the parameters a, $b_1, \ldots, b_M$, c, and d, Gneiting et al. (2005) employ minimum score or minimum contrast estimation: They choose a proper scoring rule, write the score for the training data as a function of the parameters to estimate and numerically minimise this function. The results can be interpreted as the coefficients for which this particular score would have been minimal, when summed up over the training data. Picking a proper scoring rule ensures that calibration and sharpness are addressed simultaneously (Section 2.3).

In this context, the CRPS proves suitable, and Gneiting and Raftery (2007) have shown that it is more robust than e.g. the logarithmic score. Therefore, this estimation method is referred to as minimum CRPS estimation. Modelling Y with a normal distribution has the advantage of simplifying the computing of the CRPS, as a closed form can be obtained using repeated partial integration:

$$\operatorname{crps}\left(\mathcal{N}\left(\mu, \sigma^2\right), y\right) = \sigma \left\{ \frac{y-\mu}{\sigma} \left[2\,\Phi\left(\frac{y-\mu}{\sigma}\right) - 1 \right] + 2\,\varphi\left(\frac{y-\mu}{\sigma}\right) - \frac{1}{\sqrt{\pi}} \right\},$$

where $\Phi(\cdot)$ and $\varphi(\cdot)$ denote the CDF and the PDF of the standard normal distribution, respectively.

The variance coefficients c and d should be non-negative to make certain that the variance is positive and we have a valid probability distribution. Gneiting et al. (2005) set $d = \delta^2$, optimise the CRPS over δ and in this way enforce the non-negativity; here, it is not an issue for the parameter c, but generally one can proceed in the same manner as for d.

When expressing the CRPS in terms of the parameters, the average score over all k pairs of forecasts and observations contained in the training data set is noted,

$$\Gamma\left(a; b_1, \ldots, b_M; c; \delta\right) = \frac{1}{k} \sum_{i=1}^{k} \sqrt{c + \delta^2\, S_i^2} \left\{ M_i \left[2\,\Phi\left(M_i\right) - 1 \right] + 2\,\varphi\left(M_i\right) - \frac{1}{\sqrt{\pi}} \right\},$$

where

$$M_i = \frac{Y_i - (a + b_1 X_{i1} + \ldots b_M X_{iM})}{\sqrt{c + \delta^2 S_i^2}}$$

is the standardised forecast error for the ith forecast in the training data set. Using the Broyden-Fletcher-Goldfarb-Shanno algorithm, as it is implemented in the `R` function `optim`, the authors determine the minimum of the CRPS numerically. Reaching a global minimum can not be ensured, so the initial values have to be chosen carefully. The best way to do so is to take starting values based on past experiences, e.g. the estimated parameters from the day before.

3.2.1 EMOS$^+$

In the previously described method for estimating the EMOS parameters, no constraints were assumed on the ensemble member coefficients $b_1, \ldots, b_M$. However, in the context of ensemble forecasts and statistical postprocessing, it is hard to explain negative values, as these parameters can be interpreted as weights, reflecting the relative usefulness of the particular ensemble member model. Such negative coefficients might be caused by highly correlated ensemble forecasts, e.g. because the initial and boundary conditions for the underlying numerical models are provided by the same operational weather centre.

To solve this issue, Gneiting et al. (2005) proposed an extended technique, called EMOS$^+$, producing only non-negative EMOS weights. At first, the usual estimation is executed, with no constraint on $b_1, \ldots, b_M$. Then all ensemble member forecasts with negative coefficients are removed from the training data and their respective coefficients are set to zero. The ensemble mean $\overline{X}$ and variance S^2 need to be recomputed, both for the training data and the forecast, using only the remaining ensemble members. This process is iterated until there are only non-negative regression parameters left.

The described approach can easily be interpreted as a form of model selection, where the most useful of the ensemble member models are chosen. Highly correlated models are often eliminated, and only the most skillful of them is left in the regression equation. Usually, EMOS$^+$ yields similar results in comparison to EMOS, but improves interpretability.

3.3 EMOS for Wind Speed

The standard EMOS method, however, is not appropriate for any weather variable of interest. Unlike temperature or air pressure, wind speed can not be described with a normal distribution, as it is a non-negative quantity. Therefore, Thorarinsdottir and Gneiting (2010) suggested the use of a truncated normal distribution with a cut-off at zero,

$$Y \mid X_1, \ldots, X_M \sim \mathcal{N}^0\left(\mu, \sigma^2\right).$$

The modelling of the distribution characteristics stays unchanged, the mean is a linear function of the ensemble member forecasts,

$$\mu = a + b_1 X_1 + \ldots + b_M X_M,$$

and the variance an affine function of the ensemble spread,

$$\sigma^2 = c + d\, S^2.$$

For a truncated normal distribution, the density function can be expressed in terms of the density and distribution functions of a standard normal distribution, φ and Φ:

$$f^0(y) = \frac{1}{\sigma}\, \varphi\left(\frac{y-\mu}{\sigma}\right) \Big/ \Phi\left(\frac{\mu}{\sigma}\right)$$

when $y > 0$ and $f^0(y) = 0$ otherwise.

As before, the authors restrict the spread parameters to be non-negative, just as the ensemble member coefficients $b_1, \ldots, b_M$. In this context, however, they write $b_1 = \beta_1^2, \ldots, b_M = \beta_M^2$, $c = \gamma^2$, and $d = \delta^2$, instead of employing EMOS$^+$. The parameters a, $\beta_1, \ldots, \beta_M$, γ, and δ are unconstrained real values and are again obtained by minimum CRPS estimation over a training data set. The CRPS, as defined in Section 2.3, has a closed form for the truncated normal distribution; expressed in terms of the training data and as a function of the parameters to estimate, it becomes

$$\Gamma\left(a;\beta_1,\ldots,\beta_M;\gamma;\delta\right) = \frac{1}{k}\sum_{i=1}^{k}\Phi\left(N_i\right)^{-2}\sqrt{\gamma^2+\delta^2S_i^2}\left\{M_i\,\Phi\left(N_i\right)\left[2\,\Phi\left(M_i\right)+\Phi\left(N_i\right)-2\right]\right.$$
$$\left.+\,2\,\varphi\left(M_i\right)\Phi\left(N_i\right)-\frac{1}{\sqrt{\pi}}\,\Phi\left(\sqrt{2}\,N_i\right)\right\},$$

where

$$M_i=\frac{y_i-(a+\beta_1^2X_{i1}+\ldots+\beta_M^2X_{iM})}{\sqrt{\gamma^2+\delta^2S^2}} \quad \text{and} \quad N_i=\frac{a+\beta_1^2X_{i1}+\ldots+\beta_M^2X_{iM}}{\sqrt{\gamma^2+\delta^2S^2}}.$$

Once more, the actual numerical optimisation, i.e. finding the parameter values which minimise the CRPS for the training data, rests on the Broyden-Fletcher-Goldfarb-Shanno algorithm. The initial values for this algorithm are the estimated parameters of the previous day, to enhance the probability of finding a global minimum.

3.4 EMOS for Wind Gust

The forecasting of wind gust is challenging, as it is typically not an output of numerical weather prediction models. There seems to be a direct relation to wind speed, though, namely through a gust factor τ, the ratio between the maximum gust speed and the sustained wind speed. Thorarinsdottir and Johnson (2011) developed a forecasting method depending on the previously described EMOS method for wind speed. It consists of three components which have to be predicted:

1. *Daily maximum wind speed:*
   For the daily maximum wind speed, the EMOS model defined above is employed to postprocess the ensemble forecasts and obtain a full predictive distribution. In a first step, the parameters a, $b_1,\ldots,b_M$, c, and d are estimated with minimum CRPS estimation.

2. *Probability of gust:*
   The gust speed Z can be assumed to have a multiplicative relationship to the wind speed Y,
   $$Z=\tau Y, \quad \tau\geq 1.$$

Then the conditional predictive distribution for Z becomes

$$Z \mid X_1, \ldots, X_M \sim \mathcal{N}^0\left(\tau\mu, (\tau\sigma)^2\right),$$

with μ and σ^2 denoting the mean and variance estimates for the wind speed.
In the context of the article, using observations from the Automated Surface Observing System (ASOS) network (National Weather Service 1998), gust is observed if the gust speed, determined under certain conditions, is at least 14 knots. The authors introduce two separate gust factors, τ_1 for the probability of gust and τ_2 for the conditional gust speed forecast. Thus, in the former case, they draw on the predictive distribution for Z and express this probability in terms of τ_1:

$$\mathbb{P}\left(Z \geq 14 \mid X_1, \ldots, X_M\right) = \left[1 - \Phi\left(\frac{14 - \tau_1\mu}{\tau_1\sigma}\right)\right] \Big/ \Phi\left(\frac{\mu}{\sigma}\right).$$

This parameter can also be estimated by minimum score estimation, using the Brier skill score for categorical events (e.g. Gneiting and Raftery 2007). The task is to find the value for τ_1 which minimises

$$\Gamma\left(\tau_1\right) = \frac{1}{k}\sum_{i=1}^{k}\left(\left[1 - \Phi\left(\frac{14 - \tau_1\mu}{\tau_1\sigma_i}\right)\right]\Phi\left(\frac{\mu_i}{\sigma_i}\right)^{-1} - \mathbb{I}\left\{Z_i \text{ observed}\right\}\right)^2$$

over the training set, where $\mathbb{I}\left\{Z_i \text{ observed}\right\}$ is 1 if gust is observed for the ith forecast case (according to the definition in National Weather Service (1998)), and zero otherwise.

3. *Daily maximum gust speed, conditional on gust being observed:*
Subsequently, the gust speed, conditional on gust being observed, needs to be addressed. Its predictive distribution can be modelled employing a truncated normal distribution, as with the wind speed, however now the cut-off is at 14. With the second gust factor τ_2, the predictive density becomes

$$f^{14}\left(z\right) = \left[\frac{1}{\tau_2\sigma}\,\varphi\left(\frac{z - \tau_2\mu}{\tau_2\sigma}\right)\right] \Big/ \left[1 - \Phi\left(\frac{14 - \tau_2\mu}{\tau_2\sigma}\right)\right]$$

for $z \geq 14$ and $f^{14}(z) = 0$ otherwise. Following the usual technique, CRPS minimisation is applied over training data to find the optimal value for τ_2. For this distribution, the CRPS can be determined as

$$\begin{aligned}
&\operatorname{crps}\left(\mathcal{N}^{14}(\tau_2\mu, \tau_2\sigma), z\right) = \\
&\quad \tau_2\sigma\,\lambda(\tau_2\mu, \tau_2\sigma)^{-2}\left\{\frac{z-\tau_2\mu}{\tau_2\sigma}\,\lambda(\tau_2\mu, \tau_2\sigma)\left[2\,\Phi\left(\frac{z-\tau_2\mu}{\tau_2\sigma}\right) + \lambda(\tau_2\mu, \tau_2\sigma) - 2\right]\right. \\
&\quad \left. +2\,\varphi\left(\frac{z-\tau_2\mu}{\tau_2\sigma}\right)\lambda(\tau_2\mu, \tau_2\sigma) - \frac{1}{\sqrt{\pi}}\left[1 - \Phi\left(\sqrt{2}\,\frac{14-\tau_2\mu}{\tau_2\sigma}\right)\right]\right\},
\end{aligned}$$

where $\lambda(a, b) = 1 - \Phi\left(\frac{14-a}{b}\right)$ and $z \geq 14$. Here, the observations z are the gust speed observations, and the parameters μ and σ originate from the estimation of the wind speed forecast.

All numerical optimisation is done using the Broyden-Fletcher-Goldfarb-Shanno algorithm. With this method, it is possible to obtain full predictive densities for a weather variable which normally is not included in numerical weather prediction models. It can, however, also be applied as a statistical postprocessing method for ensemble forecasts of gust speed.

The EMOS approach was developed to produce calibrated and sharp predictive distributions from underdispersed discrete ensemble forecasts. It has proved to be a very versatile method, being applicable to several weather variables such as temperature, air pressure, wind speed and wind gust. To adapt to local circumstances, the normal distribution can be substituted by a more adequate distribution, e.g. the Student's t distribution.

In the next chapter we will propose a bivariate variant that incorporates the correlation between two components into the model. More information about the different univariate EMOS methods, as well as results in form of case studies, can be found in the respective articles.

Chapter 4

Extending EMOS to Wind Vector Forecasts

In this chapter, a new method based on the conventional univariate EMOS postprocessing technique is designed, where ensemble forecasts of two weather quantities are addressed jointly. EMOS is a very useful and easy to implement way to improve the performance of wind speed forecasts (Thorarinsdottir and Gneiting 2010); there is also a variant of Bayesian Model Averaging for the direction of surface wind, developed by Bao et al. (2010). It is quite conceivable that especially EMOS can be applied to the latitudinal and longitudinal wind components individually. In this case, however, we do not make use of dependencies between the components. Therefore, we will introduce a method that employs a bivariate joint distribution and thus takes account of the correlation structure.

At first we have to become accustomed to the different properties of wind vectors, before we move on to examining some data and decide which bivariate distribution is best used. In the then following sections, it will be established how the model characteristics can be written in terms of the ensemble forecasts. Compared to the univariate EMOS technique, we will make some changes to the process of parameter estimation in that we will split it in three parts and use historic data to estimate the correlation structure. Still, the essential ideas stay the same: EMOS for wind vectors is based on linear regression, takes account of the spread-error relationship contained in the ensemble and produces calibrated and sharp forecasts.

4.1 Properties of Wind Vectors

Before we can define a model for the bivariate predictive distribution of wind vectors, we have to take a look at their properties and how they are measured. For this, consider Figure 4.1, where the green arrow represents the surface wind vector. It is the sum of a zonal and a meridional component, respectively named U and V. They therefore correspond to the parts of the wind flowing in west/east and north/south direction. For the creation of weather maps, the vectors are often illustrated as wind barbs, using shafts of a fixed length with different kinds of barbs and flags for different wind speeds.

Wind vectors indicate wind speed and wind direction simultaneously through the length and the angle of the vector. While the angle θ is measured in degrees (clockwise from north), the wind is always named after the direction where it originates from, so for example an angle of 90° means easterly wind, 225° southwesterly wind and so on. The wind speed w can be computed from the wind components U and V, using the euclidean norm:

$$w = \sqrt{U^2 + V^2}.$$

Surface wind is typically quantified in meters per second or in knots, where 1 knot approximately equals 0.514 $^{\text{m}}/_{\text{s}}$. Our new method will be applied to instantaneous 10 m wind, which means that the wind observations are averaged over the last 2 minutes of the hour and measured at 10 m above ground (see also Thorarinsdottir and Gneiting (2010), where the daily maximum wind speed was used). In practice, instead of the wind components, the wind direction θ and wind speed w are observed and then transformed into vectors.

Next, we will develop a two-dimensional model for the joint distribution of U and V. Our goal is to take advantage of a possible correlation between the two components and so capture dependencies and account for local terrain characteristics. We begin with a short data analysis where we use ensemble forecasts from the UWME ensemble for the Pacific Northwest of the USA and parts of Canada, as well as the corresponding verifying observations. Further information on the data can be found in Section 5.1.

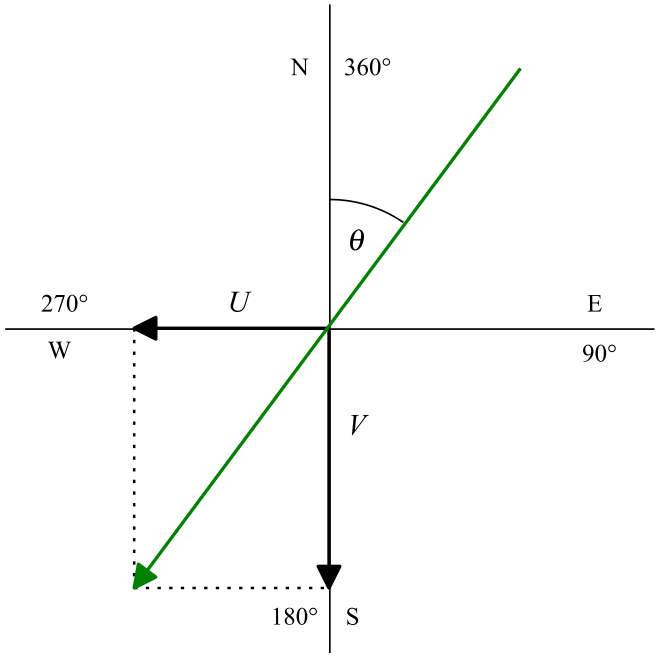


Figure 4.1: Wind vector, wind components U and V and wind direction θ

4.2 EMOS for Wind Vectors

To modify the conventional EMOS postprocessing method so that it can be applied to wind vectors, we begin by finding a bivariate distribution which represents the empirical distribution of wind vector observations as closely as possible and which is easy to work with. Again, we look at the conditional distribution of the observations given the forecasts.

In the two-dimensional case, we at first restrict the following analysis to the ensemble mean forecast, as compared to the full ensemble in the univariate case. This approach naturally applies only to ensembles with exchangeable members, but is done here because of simplicity reasons. Exchangeable members are ensemble member forecasts that only differ in random perturbations, and are therefore indistinguishable. The UWME ensemble does not consist of exchangeable members, but we will see that the performance of the new method is not very sensitive to the use of the ensemble mean only. On the contrary, it reduces the amount of parameters to estimate and helps simplifying the model.

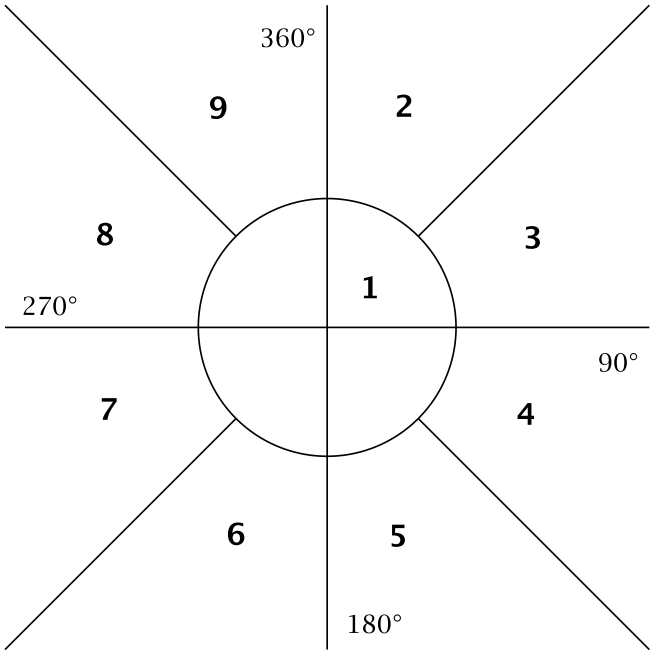


Figure 4.2: The $\mathbb{R}^2$ plane, divided into nine sections

4.2.1 Data Analysis

In the univariate EMOS method, it was established that the predictive distribution of the weather quantity of interest, conditional on the ensemble forecasts, can be modeled with a normal distribution. For the two-dimensional wind vectors, we hope to employ a bivariate normal distribution, but we first have to test if this conjecture is justified.

To examine the wind vector distribution empirically, we use data provided by the UWME, consisting of the ensemble forecasts and the corresponding verifying observations. For this purpose, we divide the plane of real numbers into nine sections (according to Figure 4.2) and, for each section, we plot the observations conditional on the corresponding ensemble mean forecast falling into this section. So, for the predictive distribution to be bivariate normal, these scatterplots should roughly have the shape of a circle or an ellipse, depending on the correlation between the vector components. To create these plots, data from the years 2007 and 2008 at 79 stations is used. All values are given in $^{\mathrm{m}}/_{\mathrm{s}}$.

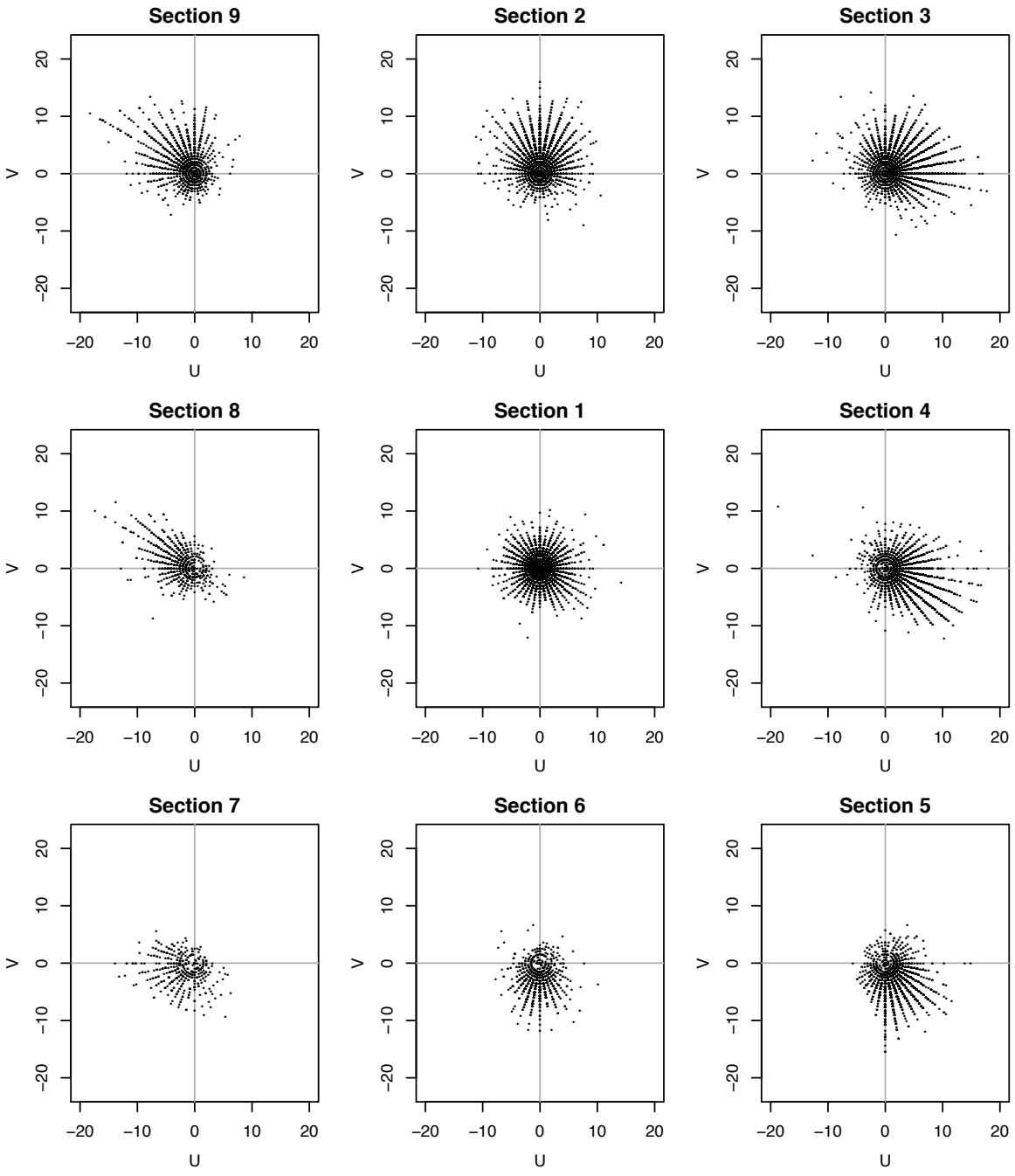


Figure 4.3: Scatterplot of the distribution of wind vector observations over the Pacific Northwest in 2007, conditional on the mean of the UWME forecast falling in the same section of the plane; the section division and arrangement are given in Figure 4.2

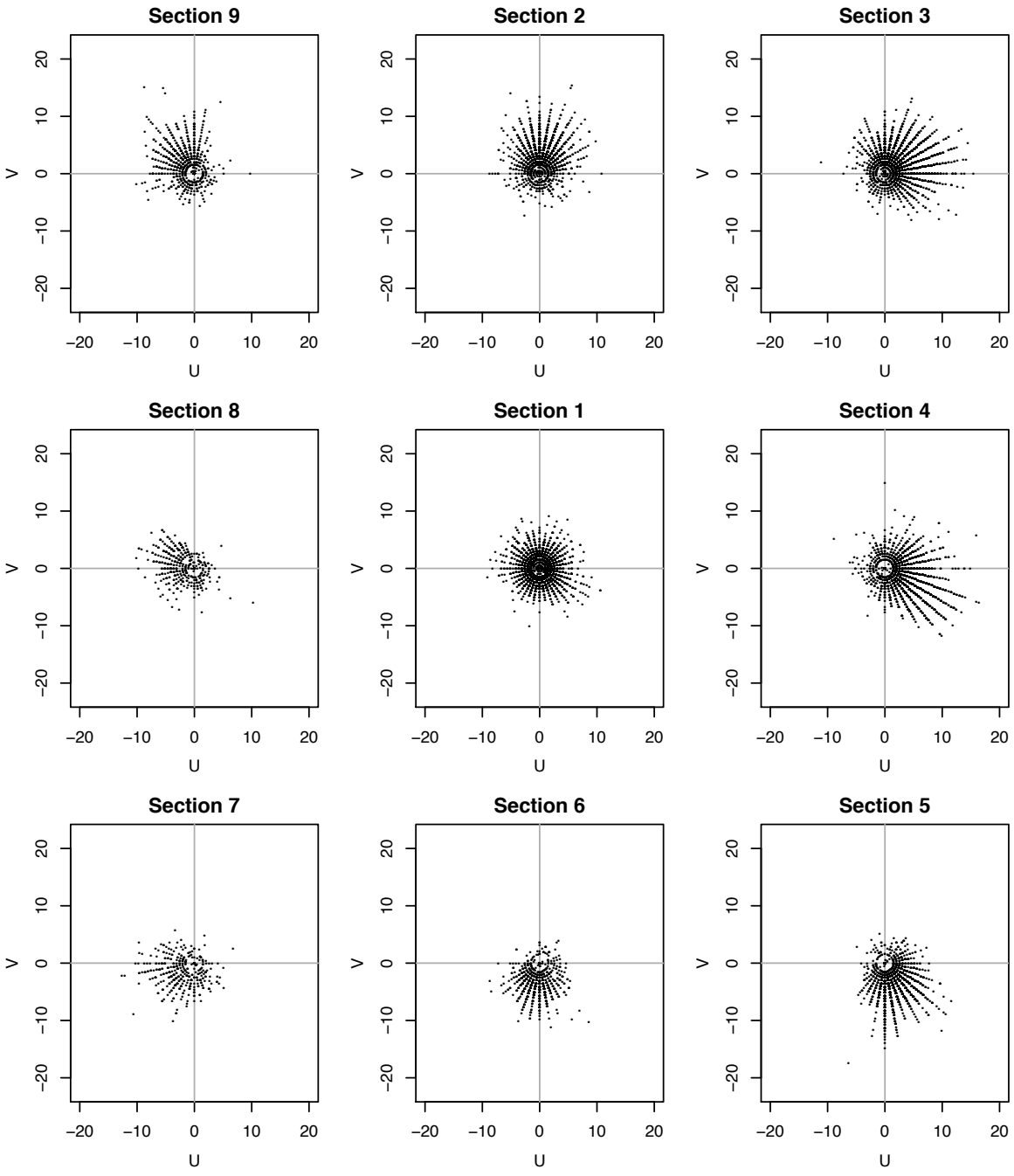


Figure 4.4: Scatterplot of the distribution of wind vector observations over the Pacific Northwest in 2008, conditional on the mean of the UWME forecast falling in the same section of the plane; the section division and arrangement are given in Figure 4.2

The circle surrounding section 1 has a radius of 2 $^{m}/_{s}$ and the numbers of the other sections are assigned clockwise, starting at 360°. In Figure 4.3 it can be seen that the observations indeed seem to be bivariate normally distributed, with the majority of points in the centre of each plot, decreasing towards the outer edges. Especially the plots for sections 4, 5, 8 and 9 show a substantial correlation between the two wind components in northwestern/southeastern direction.

If we compare the plots from the years 2007 and 2008 (Figures 4.3 and 4.4), it is obvious that the wind patterns don't seem to change very much over the years, which we will use later when estimating the predictive correlation coefficient in Section 4.2.4. Also, it can be noted that the observational data has been discretised due to the rounding of wind speed to whole knots (compare Section 5.1).

4.2.2 Modelling the predictive distribution

In this section, we will determine how the characteristics of the predictive distribution can be modelled with regard to the univariate EMOS method. The bivariate normal distribution has five characteristic values:

- the mean value for each component, μ_U and μ_V
- the variance for each component, σ_U^2 and σ_V^2
- the correlation coefficient between the two components, ρ

As we will make forecasts for all observational stations and every single day, these values depend on the location s and the time t:

- $\mu_U(s,t)$, $\mu_V(s,t)$
- $\sigma_U^2(s,t)$, $\sigma_V^2(s,t)$
- $\rho(s,t)$

We then denote the wind vector distribution, conditional on the ensemble mean vector, as

$$\begin{aligned}\begin{pmatrix} U(s,t) \\ V(s,t) \end{pmatrix} &\sim \mathcal{N}_2\left(\boldsymbol{\mu}(s,t), \boldsymbol{\Sigma}(s,t)\right) \\ &= \mathcal{N}_2\left(\begin{pmatrix} \mu_U(s,t) \\ \mu_V(s,t) \end{pmatrix}, \begin{pmatrix} \sigma_U^2(s,t) & \rho(s,t)\,\sigma_U(s,t)\,\sigma_V(s,t) \\ \rho(s,t)\,\sigma_U(s,t)\,\sigma_V(s,t) & \sigma_V^2(s,t) \end{pmatrix}\right),\end{aligned}$$

which means that the vector of wind observations at location s and time t is bivariate normally distributed with the mean vector $\boldsymbol{\mu}(s,t)$ and variance-covariance matrix $\boldsymbol{\Sigma}(s,t)$ that are valid for this time and location.

Now, we need to describe these distribution characteristics in terms of the ensemble forecasts. If we follow the method developed in Gneiting et al. (2005) and Thorarinsdottir and Gneiting (2010) closely, we have to employ a proper scoring rule, which we minimise to estimate the parameters of the predictive distribution. The preferred scoring rule so far has been the CRPS, which is a robust score and has a closed form for the normal distribution. A natural generalisation to multiple dimensions is the energy score, so a straightforward way would be to optimise this score for the predictive distribution defined above.

However, as there is no integral representation for the energy score but only one that uses the expected value of the distribution, we are not aware of a closed form for this score and minimisation could only be done numerically. This would lead to an immense increase of computational costs for this otherwise very parsimonious method (compared to BMA for example, where a mixture of several probability densities is used; see Raftery et al. (2005)).

Instead, we will separate the estimation process into three parts and use three different methods to obtain the parameters. We will use linear regression for the mean vector and maximum likelihood estimation for the variances. Finding a way to compute the correlation coefficient is challenging, as it strongly depends on the wind direction, which again follows certain patterns for different locations.

Here we will introduce a technique where a specific correlation function is estimated from the previous year's observations and is used for the forecasts of a whole year. Later we will establish that this rather complex way of determining ρ is worth the effort compared to a

naive approach where we estimate the two wind components separately with no correlation at all (Section 5.3.3).

The eventual parameter estimation will therefore be executed following these three steps:

1. Estimate the predictive means by performing linear regression on the ensemble means.

2. Compute the predictive correlation of the wind components with help of the previously estimated correlation function and the predictive ensemble means from the first step.

3. Insert the results from steps 1 and 2 into the log-likelihood function for the predictive distribution and estimate the remaining variances.

In the following sections, the modelling of the predictive distribution characteristics and the different ways of estimation will be presented.

4.2.3 The Mean Vector and Linear Regression

The analysis of the UWME data yielded that the distribution of wind vector observations conditional on the ensemble mean is approximately bivariate normal. Therefore we suggest the ensemble mean as sole predictor for the EMOS mean, in order to reduce the number of parameters to estimate and to avoid a possible overparameterisation. Of course, this supposition again only holds for ensembles with exchangeable members, which is not the case for the UWME. This will be discussed further in Section 5.2.1.1 and we will examine if a

Table 4.1: Estimated parameters for the predictive mean at station KSEA valid on 23 January 2008

a_U	b_U	a_V	b_V
-0.69	0.86	-0.42	1.08

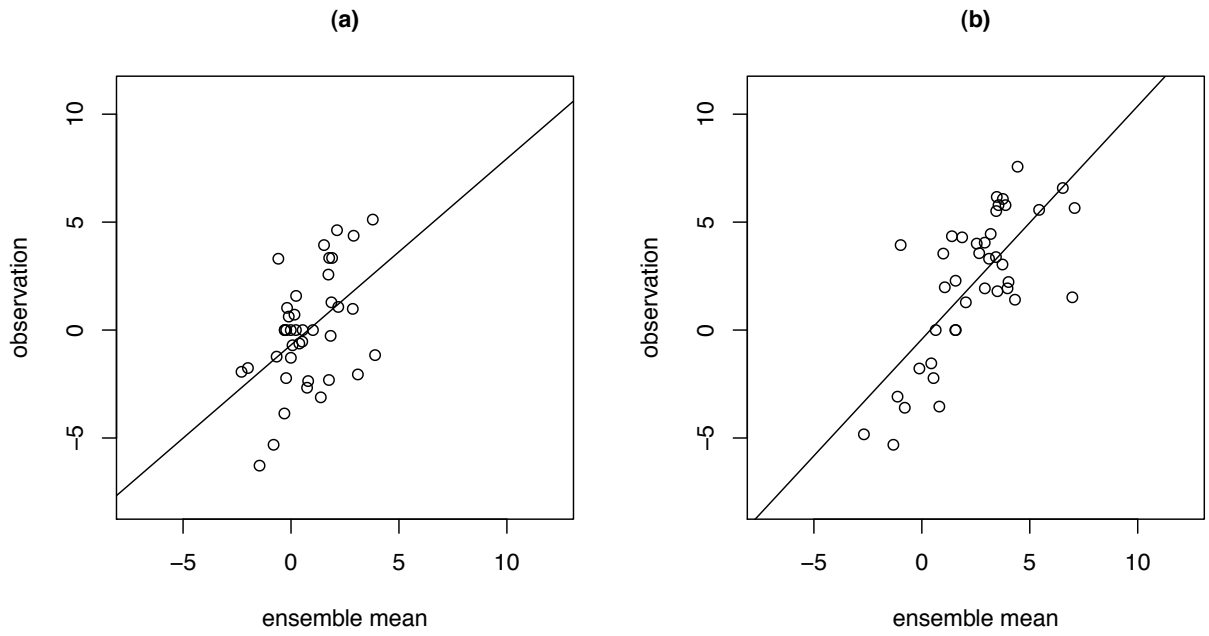


Figure 4.5: Parameter estimation of the predictive mean vector for the 48-hour-ahead forecasts at Sea-Tac Airport, valid on 23 January 2008; pictured are the scatterplots of the ensemble means and the verifying observations contained in the training data set, as well as the respective fitted linear functions, for (a) the U component and (b) the V component

model using multiple linear regression with the initial ensemble forecasts as predictors shows better results for this ensemble.

So, with respect to the univariate method, we write the predictive means as linear functions of the ensemble means $\overline{U}(s,t) = \frac{1}{M}\sum_{i=1}^{M} U_i(s,t)$ and $\overline{V}(s,t) = \frac{1}{M}\sum_{i=1}^{M} V_i(s,t)$:

$$\begin{aligned} \mu_U(s,t) &= a_U + b_U \cdot \overline{U}(s,t) \\ \mu_V(s,t) &= a_V + b_V \cdot \overline{V}(s,t)\,. \end{aligned}$$

Because only the four parameters a_U, a_V, b_U, and b_V (valid at station s and time t) are determined in this step, the best way to do so is using standard linear regression, for each component separately. Therefore we assume that we have a linear relationship between the

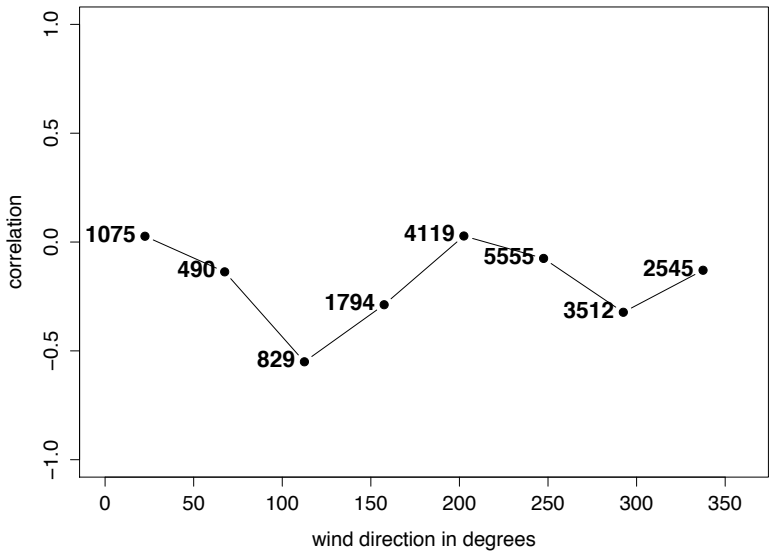


Figure 4.6: Relationship between the wind direction and the correlation based on the observations at all stations in 2007; for each section, the point represents its average wind direction while the number refers to the observation count

ensemble means and the observations and that there are error terms ε_U and ε_V, which have a univariate normal distribution with mean zero:

$$\begin{aligned} U(s,t) &= a_U + b_U \cdot \overline{U}(s,t) + \varepsilon_U \\ V(s,t) &= a_V + b_V \cdot \overline{V}(s,t) + \varepsilon_V, \end{aligned}$$

where (s,t) are the pairs of locations and dates contained in the training data. To actually compute the parameters, we use ordinary least squares, as implemented in the `R` function `lm`.

Now we take a look at an example where we make a forecast for the EMOS predictive mean vector at Sea-Tac Airport (station KSEA) on 23 January 2008. The training data for the linear regression consists of the 40 most recent pairs of ensemble means and verifying observations at this location, and the resulting estimates for the parameters are given in Table 4.1. In Figure 4.5 we can see the observations plotted over the ensemble means and the

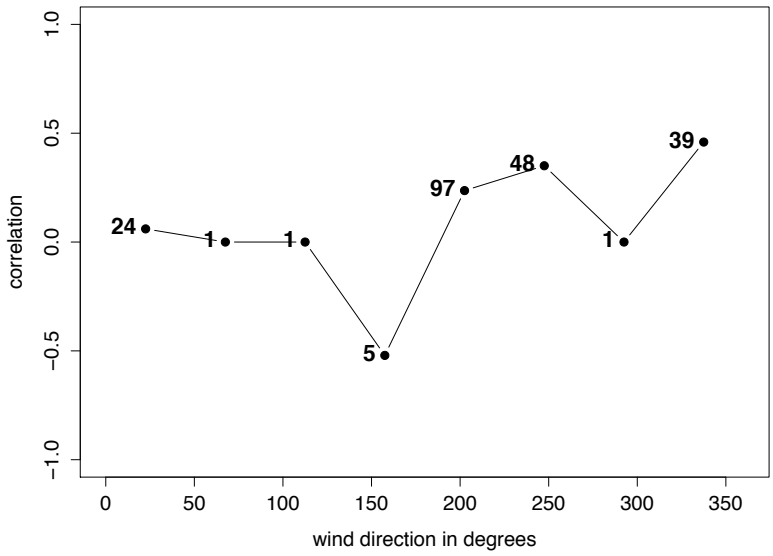


Figure 4.7: Relationship between the wind direction and the correlation based on the observations at station KSEA in 2007; for each section, the point represents its average wind direction while the number refers to the observation count

respective fitted linear functions. The ensemble mean vector for this day is $(0.51, -0.01)^T$, so the parameter estimation results in the predictive mean vector $(-0.25, -0.44)^T$.

4.2.4 The Correlation Coefficient and the Wind Direction

Our next step in the estimation process is to develop a method in order to model the correlation coefficient of the predictive normal distribution. In Section 4.2.1, we looked at scatter plots of wind vector observations in different sections of the two-dimensional plane and established that there seems to be a substantial correlation in several of them. This suggests that there is a strong connection between the direction of the forecast mean vector and the correlation of the two observation components U and V.

Further analysis of the data (Figures 4.6 and 4.7) will yield that the correlation in fact changes with a period of about 180°. The data used for this analysis is from 2007; we will also use it for the estimation of the correlation curve. For Figure 4.6, the correlation between

the observations for eight of the nine sections was computed and then plotted against the respective wind direction, with the numbers referring to the actual observation count in each section. We will not use section 1, the circle around zero, as it can not be associated with one single direction.

To obtain the actual predictive wind direction θ (in degrees) from the mean vector components μ_U and μ_V (as in Figure 4.1), the two-argument arctangent function is employed:

$$\begin{aligned}\theta &= \frac{360^\circ}{2\pi} \cdot \operatorname{atan2}\left(-\mu_U, -\mu_V\right) \\ &= \frac{360^\circ}{2\pi} \cdot \begin{cases} \arctan \frac{\mu_U}{\mu_V} & \mu_V < 0 \\ \arctan \frac{\mu_U}{\mu_V} + \pi & \mu_U \leq 0,\ \mu_V > 0 \\ \arctan \frac{\mu_U}{\mu_V} - \pi & \mu_U > 0,\ \mu_V < 0 \\ \frac{\pi}{2} & \mu_U < 0,\ \mu_V = 0 \\ -\frac{\pi}{2} & \mu_U > 0,\ \mu_V = 0 \\ 0 & \mu_U = \mu_V = 0. \end{cases}\end{aligned}$$

As this function ranges from -180° to 180°, we add 360° to negative values and therefore get results between 0° and 360°.

The shape of the resulting curve resembles that of a trigonometric function, so we can make a first guess for modelling the correlation. A cosine function with four parameters seems reasonable:

$$\rho(s,t) = r \cdot \cos\left(k \cdot \theta(s,t) + \varphi\right) + p,$$

where $\theta(s,t)$ is the wind direction of the predictive mean vector obtained in step 1. We have to constrain the period to be an integer so that the consistency between the values at e.g. 0° and 360° is retained, thus k has to be an integral multiple of $\frac{2\pi}{360^\circ}$. Figure 4.6 suggests $k = 2 \cdot \frac{2\pi}{360^\circ}$. Furthermore, considering the same plot but with data from only one station, Figure 4.7 shows that the wind direction/correlation relationship can vary locally. Due to lack of data, this station-specific curve however potentially changes over time.

An appropriate way to estimate the parameters r, φ, and p is non-linear least squares regression, while k has to be determined empirically because of the constraint mentioned

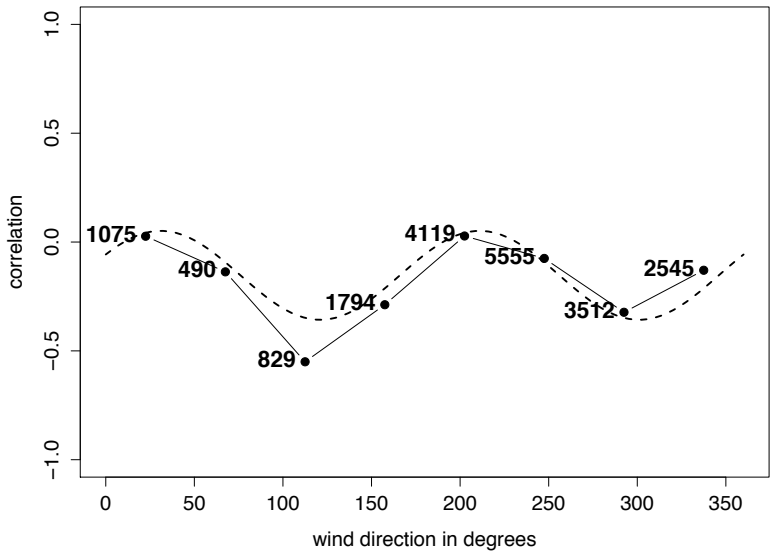


Figure 4.8: Empirical (solid line) and estimated (dashed line) correlation functions, based on all available observations in the Pacific Northwest in 2007

above. A special wind pattern at a certain location can cause few or no observations in a section, resulting in correlations of 1 or -1. The number of observations in each section will serve as weights, therefore taking account of this unreasonable extremal values when there is only little data available for a certain direction.

So, for estimating this trigonometric correlation function, we use the built-in `R` function `nls` with the relative frequency of observations as weights. If we use the data from all stations, we get

$$\rho(s,t) = 0.20 \cdot \cos\left(\frac{4\pi}{360^\circ} \cdot \theta(s,t) - 1.08\right) - 0.15,$$

which is shown in Figure 4.8. For the example where data from only one station is used, the weighted regression yields the curve in Figure 4.9; we chose a period of 360°, that is $k = \frac{2\pi}{360^\circ}$. The individual empirical and estimated correlation curves for all 79 stations and the respective choices for the parameter k can be found in Appendix A.

With this function we have found a way to describe and measure the relationship between the wind direction and the correlation of the predictive distribution. It was established earlier

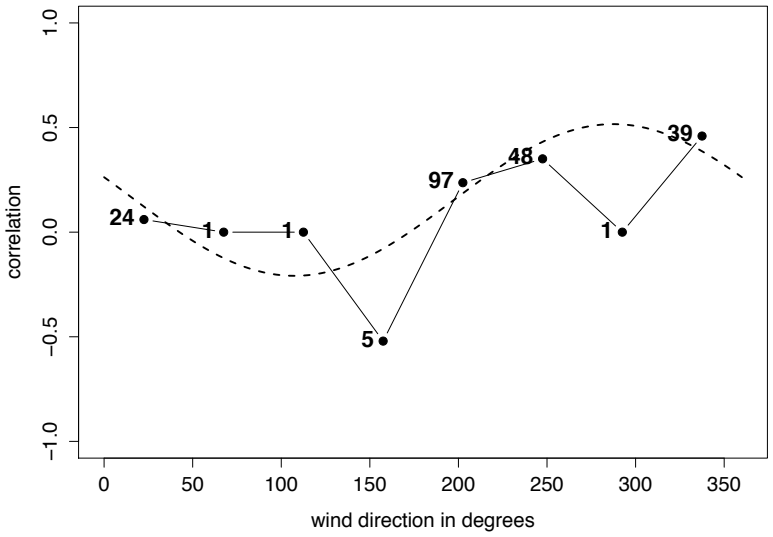


Figure 4.9: Empirical (solid line) and estimated (dashed line) correlation functions, based on all available observations at Sea-Tac Airport in 2007

that the specific wind patterns don't seem to vary much from year to year, so that we will use the curves estimated with the 2007 data for issuing out-of-sample predictions in 2008. In continuing the example from the previous section, we take the bias-corrected predictive mean $(-0.25, -0.44)^T$, compute the predicted wind direction $\theta = 30.24°$ and with this the predictive correlation $\rho = 0.07$. Here we use the location-specific correlation function for Sea-Tac Airport.

This brings the characterization of step 2 of the parameter estimation to an end, all that remains is to employ the maximum likelihood technique and estimate the predictive variances σ_U^2 and σ_V^2.

4.2.5 The Variances and Maximum Likelihood

In this section, we return to the conventional EMOS method and optimise a proper scoring rule to estimate the remaining parameters. Also, we want to model the predictive variances σ_U^2 and σ_V^2 as affine functions of the ensemble variances according to the one-dimensional variant:

$$\begin{aligned}\sigma_U^2(s,t) &= c_U + d_U \cdot S_U^2(s,t) \\ \sigma_V^2(s,t) &= c_V + d_V \cdot S_V^2(s,t)\,,\end{aligned}$$

where $S_U^2(s,t) = \frac{1}{M}\sum_{i=1}^{M}\left(U_i(s,t) - \overline{U}(s,t)\right)^2$ and $S_V^2(s,t) = \frac{1}{M}\sum_{i=1}^{M}\left(V_i(s,t) - \overline{V}(s,t)\right)^2$ are the respective variances of the ensemble forecasts at location s and time t. These quantities can be interpreted as the variances of the error terms ε_U and ε_V introduced in Section 4.2.3 and they take account of the relationship between the ensemble spread and the forecast error often observed in ensemble forecasts.

The parameters c_U, c_V, d_U, and d_V need to be non-negative to ensure that the variances are always positive and that by applying this method the spread is enlarged, as we wish to correct the underdispersion of the ensemble. Thus we employ the approach suggested in Thorarinsdottir and Gneiting (2010), writing $c_U = \gamma_U^2$, $c_V = \gamma_V^2$, $d_U = \delta_U^2$, $d_V = \delta_V^2$, and

$$\begin{aligned}\sigma_U^2(s,t) &= \gamma_U^2 + \delta_U^2 \cdot S_U^2(s,t) \\ \sigma_V^2(s,t) &= \gamma_V^2 + \delta_V^2 \cdot S_V^2(s,t)\,.\end{aligned}$$

To obtain these parameters, we will use the classical method of maximum likelihood estimation. It corresponds to finding the values for the parameters under which the probability that the training data would have been observed is maximal. This probability is expressed by the likelihood function

$$L(\boldsymbol{\eta}) = \prod_{(s,t)} f(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\eta})$$

with $f(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\eta})$ being the density of the bivariate normal predictive distribution and $\boldsymbol{\eta} = (\gamma_U, \gamma_V, \delta_U, \delta_V)^T$ the vector of parameters to estimate. Here $\mathbf{x} = (U, V)^T$ is the vector of verifying observations, $\boldsymbol{\mu}$ the mean vector and $\boldsymbol{\Sigma}$ the variance-covariance matrix for all

Table 4.2: Estimated parameters for the predictive variance at station KSEA valid on 23 January 2008

c_U	d_U	c_V	d_V
2.29	2.67	5.01	0.00

forecasts contained in the training set (s,t). In practice it is often easier to maximise the logarithm of the likelihood function, the log-likelihood function $l(\boldsymbol{\eta})$, for reasons of algebraic simplicity and numerical stability (Raftery et al. 2005). Thus our function to optimise becomes

$$
\begin{aligned}
l(\boldsymbol{\eta}) &= \log\left(\prod_{(s,t)} f(\mathbf{x},\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\eta})\right) \\
&= \log\left(\prod_{(s,t)} \frac{1}{2\pi\sqrt{\det\boldsymbol{\Sigma}}}\cdot\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\mathbf{x}-\boldsymbol{\mu})\right)\right) \\
&= \sum_{(s,t)} \log\left[\frac{1}{2\pi\sqrt{\det\Sigma}}\cdot\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\mathbf{x}-\boldsymbol{\mu})\right)\right] \\
&= \sum_{(s,t)} \left[\log\left(\frac{1}{2\pi\sqrt{\det\boldsymbol{\Sigma}}}\right) - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\mathbf{x}-\boldsymbol{\mu})\right] \\
&= \sum_{(s,t)} \left[\log\left(\frac{1}{2\pi\sqrt{\sigma_U^2\sigma_V^2(1-\rho^2)}}\right) - \frac{1}{2}\begin{pmatrix} U-\mu_U \\ V-\mu_V\end{pmatrix}^T\begin{pmatrix}\sigma_U^2 & \rho\sigma_U\sigma_V \\ \rho\sigma_U\sigma_V & \sigma_V^2\end{pmatrix}\begin{pmatrix} U-\mu_U \\ V-\mu_V\end{pmatrix}\right] \\
&= \sum_{(s,t)} \left[\log\left(\frac{1}{2\pi\sqrt{\gamma_U^2+\delta_U^2\cdot S_U^2(s,t)}\sqrt{\gamma_V^2+\delta_V^2\cdot S_V^2(s,t)}\sqrt{1-\rho^2(s,t)}}\right)\right. \\
&\qquad - \frac{[U(s,t)-\mu_U(s,t)]^2}{2[1-\rho^2(s,t)][\gamma_U^2+\delta_U^2\cdot S_U^2(s,t)]} \\
&\qquad - \frac{[V(s,t)-\mu_V(s,t)]^2}{2[1-\rho^2(s,t)]\,[\gamma_V^2+\delta_V^2\cdot S_V^2(s,t)]} \\
&\qquad \left. + \frac{\rho(s,t)[U(s,t)-\mu_U(s,t)][V(s,t)-\mu_V(s,t)]}{[1-\rho^2(s,t)][\gamma_U^2+\delta_U^2\cdot S_U^2(s,t)][\gamma_V^2+\delta_V^2\cdot S_V^2(s,t)]}\right].
\end{aligned}
$$

The values of the mean vector and the correlation coefficient don't depend on $\boldsymbol{\eta}$ and are previously computed in step 1 and step 2. Of course, this function can only be applied if we assume that the forecast errors are independent of time and location. This assumption usually

doesn't hold, but as in Raftery et al. (2005) and Gneiting et al. (2005), possible dependencies won't have a large effect on the predictions because we only make forecasts for one time and one location simultaneously.

The actual numerical maximisation is done with the `optim` function in `R` using the Broyden-Fletcher-Goldfarb-Shanno algorithm. Maximum likelihood estimation can be perceived as optimising the ignorance score (compare Section 2.3), so this method is equivalent to optimising a proper scoring rule, similar to the univariate case. As initial values for the algorithm, we use the results of the previous day's estimation. This method does not necessarily find a global maximum, but in this case we found the maximisation to be relatively insensitive to the choice of starting values.

For our example, Sea-Tac Airport on 23 January 2008, the estimated variance parameters are shown in Table 4.2. The ensemble variances S_U^2 and S_V^2 are 0.19 and 0.69, respectively, so the estimation results are the predictive variances $\sigma_U^2 = 2.79$ and $\sigma_V^2 = 5.01$. Larger values for d_U or d_V represent a high spread-error correlation in the training data, whereas on days where there seems to be no spread-error correlation at all, the corresponding values will be very small or even zero, like for the V component in our example.

In this manner, we conclude the description of our new EMOS postprocessing method, which is applicable to wind vectors, and proposes a way to extend the usual parameter estimation to multiple dimensions. The major changes compared to the univariate case is that we split the process into three parts and estimate the correlation coefficient with a trigonometric function of the predicted wind direction. In Section 5.4 we will see that this procedure results in a better performance than other approaches such as error dressing, separate postprocessing for the individual wind vector components or the Ensemble Copula Coupling approach, where we take account of the dependency structure of the ensemble.

Chapter 5

Case Study and Predictive Performance

After developing a bivariate EMOS method in the previous chapter, we now analyse its out-of-sample predictive performance by conducting a case study for wind vector forecasts over the North American Pacific Northwest. This particular ensemble prediction system was also used to test the skill of the univariate EMOS variants described in Chapter 3. We compare two-dimensional EMOS to the unprocessed ensemble and a number of other forecasts taking the form of ensembles or densities. They include a climatological forecast based on past observations, an ensemble dressing technique adding past errors to the ensemble mean, a combination of univariate EMOS predictive distributions for the individual vector components, and a multivariate postprocessing method using copulas. Finally, we examine if the bivariate method also creates calibrated and sharp wind speed forecasts and we assess its performance relative to the original EMOS method for wind speed.

5.1 The University of Washington Mesoscale Ensemble

In order to test the skill of the new EMOS postprocessing technique, we use data provided by the UWME ensemble (as described in Eckel and Mass (2005)), for the period between 1 January 2008 and 31 December 2008. During this time the eight-member multianalysis ensemble

Table 5.1: The UWME members and the sources of the different initial and lateral boundary conditions.

No	Member Name	IC / LBC Source	Operational Centre
1	AVN-MM5	Global Forecast System	USA National Centers for Environmental Prediction
2	CMCG-MM5	Global-Environmental Multi-Scale Model	Canadian Meteorological Centre
3	ETA-MM5	ETA Limited-Area Mesoscale Model	USA National Centers for Environmental Prediction
4	GASP-MM5	Global Analysis and Prediction Model	Australian Bureau of Meteorology
5	JMA-MM5	Global Spectral Model	Japanese Meteorological Agency
6	NGPS-MM5	Navy Operational Global Atmospheric Prediction System	Fleet Numerical Meteorological and Oceanographic Center
7	TCWB-MM5	Global Forecast System	Taiwan Central Weather Bureau
8	UKMO-MM5	Unified Model	UK Met Office

consisted of multiple runs of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5) with initial and lateral boundary conditions from eight different operational centres around the world. The ensemble members and the corresponding sources of initialisations are shown in Table 5.1.

The region covered by the UWME is the Pacific Northwest of the USA and Canada. An inner nest roughly consists of the US states Washington, Oregon and Idaho, as well as the southern part of the Canadian province British Columbia; the forecasts were made on a 12 km grid. The outer domain covers the western part of North America and large parts of the eastern Pacific Ocean, having a horizontal grid spacing of 36 km. We use 48-h-ahead forecasts that were issued for observational stations and bilinearly interpolated from the four

surrounding grid points. Additionally, they were rotated to match the northern direction at each station.

Observation data is provided by weather observation stations that are part of the Automated Surface Observing System network (National Weather Service 1998). Our new method is applied to horizontal instantaneous surface wind, where "instantaneous" means that the wind is measured and averaged over the last two minutes before the valid time at 00 UTC (see also Gneiting et al. (2008), where a similar data set was used). To obtain verifying wind vector observations, the observed wind direction θ and wind speed w were transformed into vectors, using the following relationships (see Figure 4.1 for reference):

$$\begin{aligned} U &= -w \cdot \sin\left(\frac{2\pi}{360^\circ} \cdot \theta\right) \\ V &= -w \cdot \cos\left(\frac{2\pi}{360^\circ} \cdot \theta\right). \end{aligned}$$

While being recorded in knots, the wind speed observations are rounded to the nearest whole knot, and values below 1 knot are recorded as 0. Therefore the wind vector observations are discretised, which is easily recognisable in Figures 4.3 and 4.4. Quality control procedures, described in Baars (2005), were applied, removing dates and locations with missing forecasts or observations from the data set.

For the time period considered, 19,282 pairs of ensemble forecasts and observations were available, corresponding to 291 forecast days. There are 79 observation locations, whose distribution is pictured in Figure 5.1. Additional data from the years 2006 and 2007 is required in order to provide an appropriate training period for all days of 2008, to estimate the predictive correlation functions and to establish the optimal length of the rolling training periods. Further information about the UWME, now using the WRF mesoscale model, as well as real time forecasts and observations can be found at `http://www.atmos.washington.edu/~ens/uwme.cgi`.

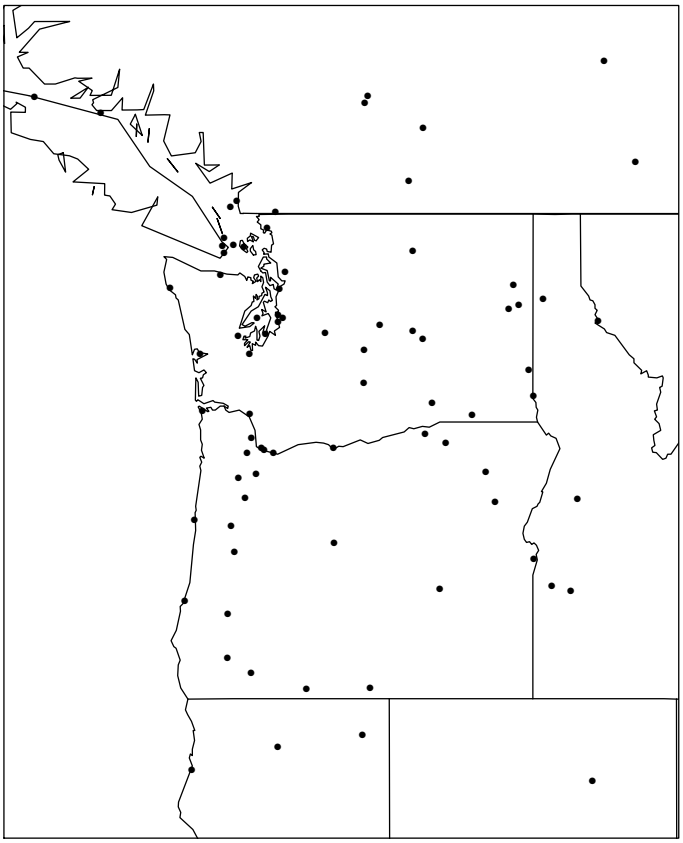


Figure 5.1: Observation locations contained in the data set

5.2 Regional and Local EMOS Technique

In concordance with the univariate EMOS technique for calibrated forecasts of wind speed (Thorarinsdottir and Gneiting 2010), we will present two techniques which differ in the choice of data to use in the training set. Furthermore, for each of these techniques, we will determine the optimal length m of the rolling training period.

Naturally, the parameters have less statistical variability when there is more data contained in the training set. However, if a long period is chosen, the model is also less adaptive to e.g. seasonal changes and varying relative performances of the underlying models. Hence, there is a trade-off between choosing the training period to be as short as possible, and making sure

that, at the same time, calibrated and sharp predictive distributions are produced. To obtain the best choice, we apply the bivariate EMOS method to data from the calendar year 2007, using additional data from 2006 to provide full training periods for all dates. This data is not used for evaluating the performance of the new EMOS method. In this experiment, we computed the out-of-sample energy score and the Euclidean error, averaged over all dates and locations, considering various training period lengths. Therefore the optimal training period length is where these scores are minimised, as proper scoring rules assess calibration and sharpness simultaneously.

5.2.1 Regional EMOS

For the regional EMOS technique, on each day only one set of parameters is estimated and used to produce forecast for all stations. The training set consists of all available data from the 79 stations and the last m days prior to the forecast being made. Should data be missing at certain dates or stations, these are omitted from the training set. For example, if we assume a training period length of 40 days, and make forecast for day 183, issued on day 181, all forecasts and observations from dates 142 to 181 are included in the training set. As for the correlation curve, all forecasts and observations of the previous calendar year, i.e. 2007, are employed to estimate one correlation function (see Figure 4.8).

This technique allows us to produce forecasts for any location in our area of interest for which ensemble forecasts are available, even if there should be no training data at this particular station. It also has the advantage of being very parsimonious, as only very few parameters have to be computed. Thus we can think about extending the parameter estimation so that we make use of the ensemble forecasts themselves.

5.2.1.1 Using Multiple Linear Regression to Estimate the Predictive Mean

As suggested before (Section 4.2.3), we consider two different kinds of estimation for the predictive mean. Initially only the ensemble mean was used as predictor for the linear regression. However, as the regional EMOS technique comes with a large amount of training data, it is possible to fit more parameters and we can also estimate the predictive mean as a linear function of the original ensemble member forecasts:

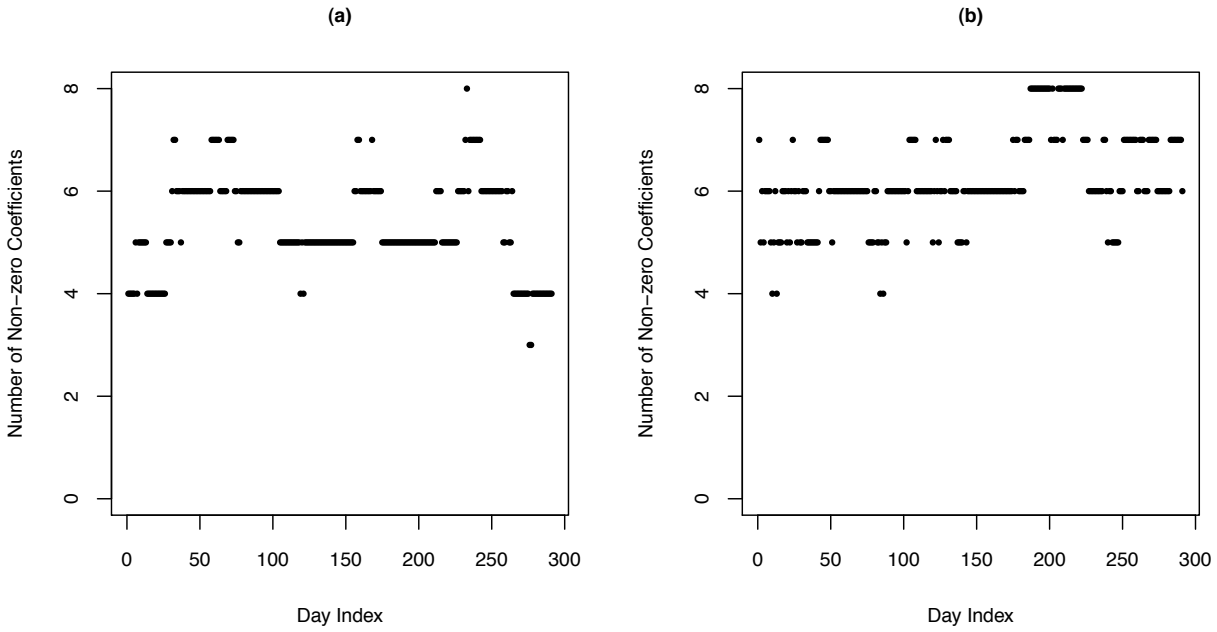


Figure 5.2: Number of non-zero EMOS$^+$ coefficients when estimating the predictive mean for all dates in the data set: (a) U wind component and (b) V wind component

$$\begin{aligned} \mu_U(s,t) &= a_U + b_{U,1} \cdot U_1(s,t) + \ldots + b_{U,8} \cdot U_8(s,t) \\ \mu_V(s,t) &= a_V + b_{V,1} \cdot V_1(s,t) + \ldots + b_{V,8} \cdot V_8(s,t)\,. \end{aligned}$$

Again, we obtain the parameters $a_U, b_{U,1}, \ldots, b_{U,8}$ and $a_V, b_{V,1}, \ldots, b_{V,8}$ by performing two separate linear regressions before the correlation coefficient and variances are estimated as usual. To improve interpretability, the coefficients $b_{U,1}, \ldots, b_{U,8}$ and $b_{V,1}, \ldots, b_{V,8}$ must be non-negative. In order to ensure this, we use a variant of the EMOS$^+$ technique proposed by Gneiting et al. (2005). As the coefficients are estimated using a linear regression model, we proceed the following way. After determining the regular regression estimates, we remove those ensemble members with non-negative parameters from the training set and set their coefficients to zero, then estimate again. This is repeated until there are only non-negative coefficients left. For the remainder of the parameter estimation, i.e. the computing of the predictive correlation and the maximum likelihood estimation, we use the ensemble variance

recomputed from the reduced ensemble. In Figure 5.2, the number of non-zero coefficients is pictured over time. We can see that on most days there are 5 or 6 ensemble members left to generate the forecast. Whether this extended method provides better results than the simple linear regression, which relies on the unjustified assumption of exchangeable ensemble members, is discussed in Section 5.4.3.1.

5.2.1.2 Length of Training Period

We established above that our choice for the size of m, which determines the amount of data contained in the training set, should be as small as possible without lack of calibration. Therefore we computed the mean energy score and Euclidean error depending on the length of the training period, using data from the year 2007, for both the simple regression and the multiple regression variant. The results are shown in Figures 5.3 and 5.4, respectively.

In the case of the simple linear regression, the variation of the training period length causes only little change in the mean value of the scores. Both the energy score and the Euclidean error reach their respective minimum around 30, which appears to be a good choice. However, any number between 20 and 40 days is acceptable.

For the multiple linear regression, more parameters have to be estimated, which is why the training set has to contain more data as for the simple regression. Looking at the mean scores in Figure 5.4 suggests a value between 40 and 50 days, and so, in order to be flexible to changing circumstances, we chose a sliding training period of 40 days.

5.2.2 Local EMOS

The local EMOS method provides a localised approach, which can be useful at sites where more specific forecasts are required, e.g. because local biases or dispersion errors need to be addressed. Here we only use training data from the station at hand and thus obtain a set of parameters for each station at every single day. The training period consists of the m most recent days where forecasts and observations were available, so that we always have m pairs of data contained in the training set. As this is significantly less data than for the regional technique, it seems not reasonable to apply multiple linear regression. In the example above, we make a forecast for day 183, issued on day 181. If no data is missing, the training period

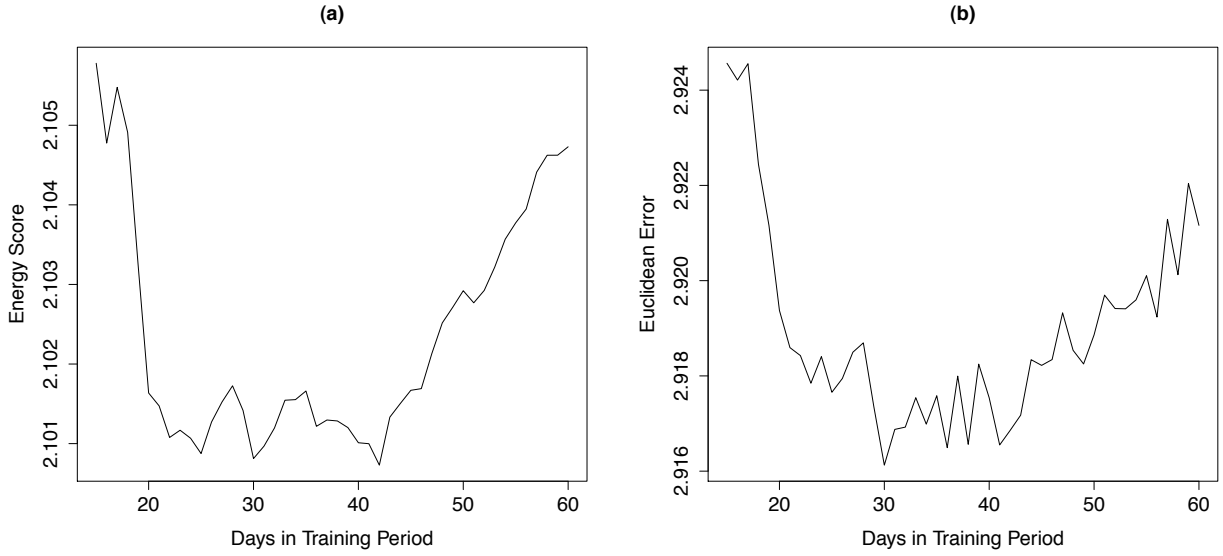


Figure 5.3: Mean energy score (a) and mean euclidean error (b) for the Pacific Northwest in 2007, as a function of the rolling training period length for the regional EMOS technique using simple linear regression

again stretches from day 142 to 181, for a length of 40 days. However, should there be 2 days with missing data at this particular station, the training set is extended and now includes all forecasts and observations for days 140 to 181.

As the correlation structure can vary locally, due to different terrains and surroundings, we produce local correlation functions for each station. They are estimated in the same manner as the regional version, but only using the location specific observations. For reference, all 79 empirical curves and the results of the estimation are shown in Appendix A.

We use the previous day's results as initial values (Section 4.2.5) for the estimation of the variance parameters. In the local approach, however, sometimes the coefficients d_U and d_V are zero, as there is no spread-error correlation for the ensemble forecasts in the training data. When this happens, we set the starting parameters δ_U and δ_V for the following day to two. This assures that, in the next step, the estimation algorithm can converge to values far from the boundary.

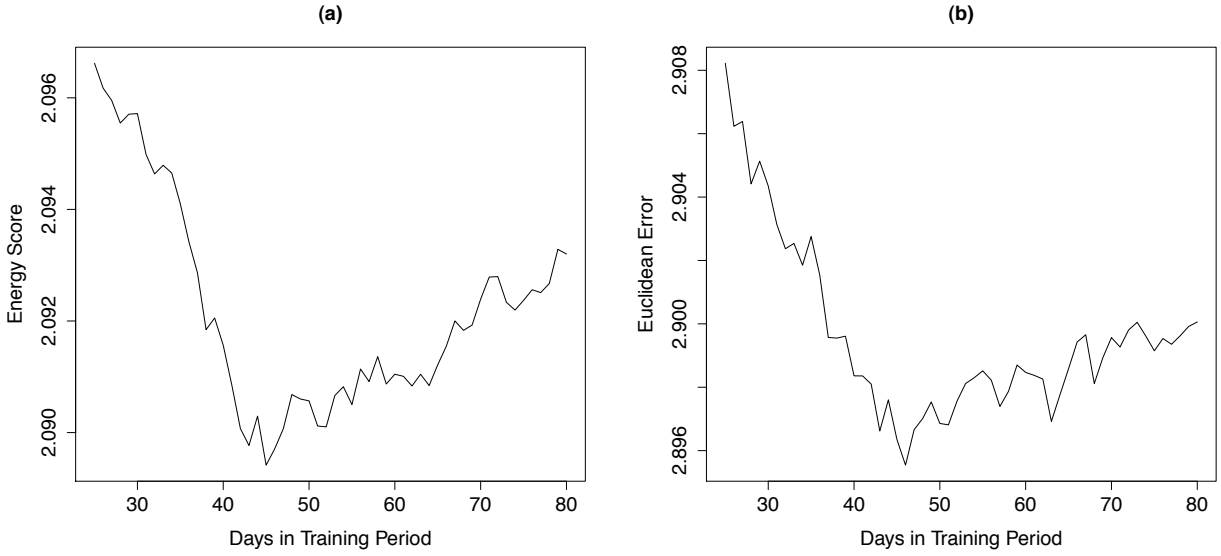


Figure 5.4: Mean energy score (a) and mean euclidean error (b) for the Pacific Northwest in 2007, as a function of the rolling training period length for the regional EMOS$^+$ technique using multiple linear regression

5.2.2.1 Length of Training Period

Following the example in Thorarinsdottir and Gneiting (2010), we only use the 30 observational locations in the US state of Washington in order to determine the local training period length and examine the mean energy score and mean euclidean error for each station. For comparability, at each station the values are divided by the respective mean over all period lengths. Figure 5.5 summarises the outcome of this analysis.

On the basis of this plot, any value between 40 and 60 days appears to be a good choice for the length of the training period. For this case study, $m = 40$ days was selected, as the training period should be as short as possible, increasing the ability to adapt to seasonal changes.

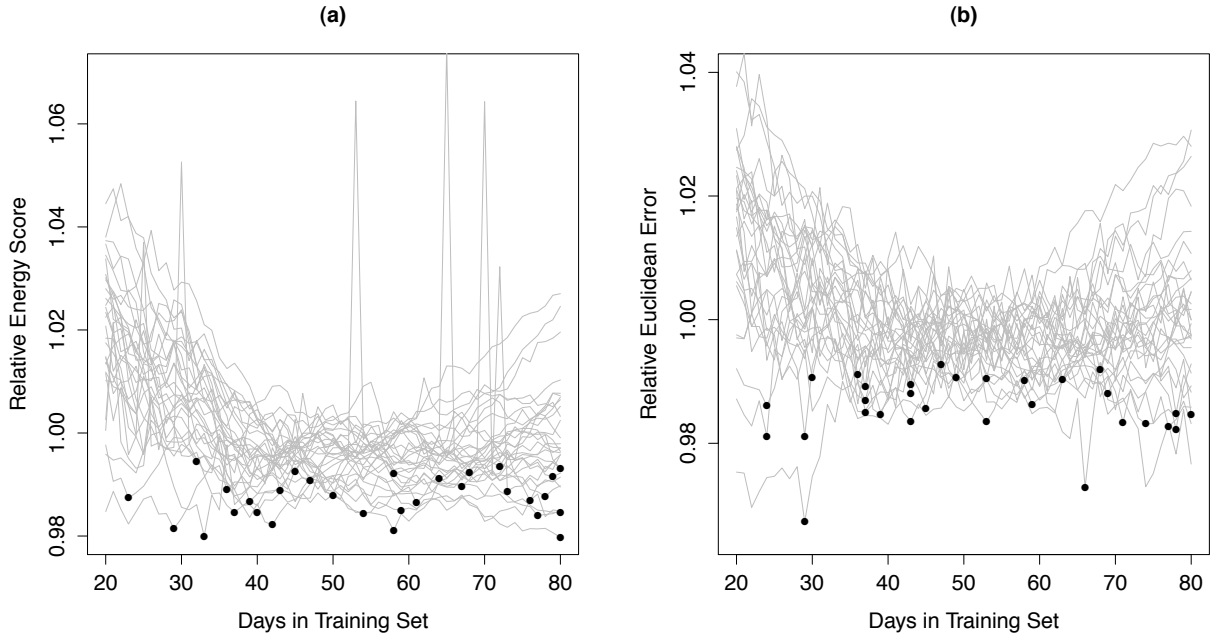


Figure 5.5: Relative mean energy score (a) and relative mean euclidean error (b) for the 30 observation stations in Washington State in 2007, as a function of the rolling training period length for the local EMOS technique; the points indicate the lowest score for each station

5.3 Competing Forecasts

Additional to EMOS and the raw ensemble, there are several other methods to predict wind vectors, which are introduced and described in this section. Their performance will be assessed and compared in Section 5.4. The climatological ensemble is a forecasting technique using weather observations, whereas the other methods are postprocessing techniques trying to obtain calibrated and sharp forecasts from the UWME ensemble.

5.3.1 Climatological Ensemble

At first, we want to introduce a climatological forecast, which means perceiving past weather observations as predictions for the future. In the context of this thesis, the climatological forecast is obtained from the training data used to determine the EMOS predictive distribution.

From the wind vector observations contained in this data, m vectors are randomly sampled, yielding an ensemble of the same size as the length of the EMOS rolling training period. Again we have a regional and a local technique, so that all forecasting methods can be compared directly. The regional technique uses observations from all locations available, and the local only from the station at hand.

As the observations are rounded and include many values equaling zero, the ensemble size is larger as compared to the UWME to avoid numerical problems on one hand and to improve the fitting of a normal distribution (as described in Section 5.3.5) on the other.

5.3.2 Error Dressing Ensemble

In Gneiting et al. (2008), a postprocessing method called error dressing was discussed, producing well-performing forecasts. It uses the raw ensemble's own statistical errors to generate a new ensemble that is corrected for biases and dispersion errors. The general approach was first proposed in Roulston and Smith (2003), where it is called "dressing" and applied to each member of the ensemble. Here we will content ourselves with creating an error dressing ensemble from the UWME mean forecast and the corresponding verifying observation. The data used are again the respective training sets for the regional and local EMOS methods.

First we compute the ensemble mean errors,

$$\begin{pmatrix} e_U(s,t) \\ e_V(s,t) \end{pmatrix} = \begin{pmatrix} U(s,t) - \overline{U}(s,t) \\ V(s,t) - \overline{V}(s,t) \end{pmatrix},$$

where (s,t) ranges over all stations and dates contained in the training data set. From this empirical error distribution, m errors $\begin{pmatrix} e_{U,1}, \ldots, e_{U,m} \\ e_{V,1}, \ldots, e_{V,m} \end{pmatrix}$ are sampled and then added to the UWME mean for the current forecast day and location. This results in the new error dressing ensemble with m ensemble members:

$$\begin{pmatrix} \overline{U} + e_{U,1} \\ \overline{V} + e_{V,1} \end{pmatrix}, \ldots, \begin{pmatrix} \overline{U} + e_{U,m} \\ \overline{V} + e_{V,m} \end{pmatrix}.$$

In contrast to Gneiting et al. (2008), we only make use of past errors to simplify comparability of all postprocessing and forecasting methods. For the same reason, both regional and local forecasts are provided.

5.3.3 EMOS for Wind Vector Components

As the wind vector data is given in form of its meridional and longitudinal components, the idea comes to mind that the conventional one-dimensional EMOS method could be applied to the two components separately. Recall that in Section 4.2.1 it was established that wind vectors seem to be normally distributed. Therefore, the resulting univariate predictive distributions can be joined to a bivariate normal distribution, assuming that U and V are independent and hence especially are not correlated.

To propose such an approach, we can follow the method as it was described in Gneiting et al. (2005) and Section 3.2 and model the predictive distribution of a wind component Y at location s and time t, conditional on ensemble forecasts $X_1, \ldots, X_M$, accordingly:

$$Y(s,t) \mid X_1(s,t), \ldots X_M(s,t) \sim \mathcal{N}\left(\mu(s,t), \sigma^2(s,t)\right).$$

The EMOS mean μ is a linear function of the ensemble members,

$$\mu(s,t) = a + b_1 X_1(s,t) + \ldots + b_M X_M(s,t),$$

and the predictive variance σ^2 an affine function of the ensemble variance S^2,

$$\sigma^2(s,t) = c + d \cdot S^2(s,t).$$

For the parameter estimation, we return to the method of CRPS minimisation and estimate all parameters simultaneously. The only difference compared to Gneiting et al. (2005) is that we do not use the method introduced as EMOS$^+$ to ensure non-negative values for $b_1, \ldots, b_M$, but rather the one by Thorarinsdottir and Gneiting (2010), where these coefficients are written as $\beta_1^2, \ldots, \beta_M^2$, additional to $c = \gamma^2$ and $d = \delta^2$. Then the estimation is performed, unconstrained, over a, $\beta_1, \ldots, \beta_M$, γ, and δ.

After the described method is applied to both U and V forecasts, we combine the results, namely μ_U, μ_V, σ_U^2, and σ_V^2, to obtain a bivariate probability distribution:

$$\begin{pmatrix} U(s,t) \\ V(s,t) \end{pmatrix} \sim \mathcal{N}_2 \left(\begin{pmatrix} \mu_U(s,t) \\ \mu_V(s,t) \end{pmatrix}, \begin{pmatrix} \sigma_U^2(s,t) & 0 \\ 0 & \sigma_V^2(s,t) \end{pmatrix} \right).$$

This distribution can be assessed in the same way as the one generated by the bivariate method taking the correlation into account. The comparison of these two distributions tells us if a two-dimensional approach to wind vectors is justified at all, or if it suffices to address the vector components individually. As before, we produce regional as well as local forecasts, using the same length of training period as in Sections 5.2.1 and 5.2.2. In the following, this method is referred to as EMOS^{UV}.

5.3.4 ECC Ensemble

The fourth technique we want to compare our EMOS method to, is called Ensemble Copula Coupling (ECC). It was developed to combine already postprocessed forecasts, so that multiple weather quantities, locations or forecast horizons can be addressed together. In this manner, multivariate dependence structures are inherited from the ensemble. More information about ECC can be found in Schefzik (2011).

In the current framework, we plan to apply this approach to the univariate EMOS^{UV} forecasts produced by the method described in the preceding section. Thus we obtain a two-dimensional ensemble that takes account of the correlation structure between the wind components, instead of just assuming that they are independent.

We proceed as follows:

1. *Employ univariate statistical postprocessing:*
   At first, we apply the aforementioned EMOS^{UV} postprocessing method for wind components to the ensemble forecasts $U_1, \ldots, U_8$ and $V_1, \ldots, V_8$. In this way, we obtain calibrated and sharp univariate predictive distributions for U and V.

2. *Draw random samples:*
   We draw samples $\hat{U}_1, \ldots, \hat{U}_8$ and $\hat{V}_1, \ldots, \hat{V}_8$ from the EMOS^{UV} predictive distributions of both U and V wind components. Here, random samples are used, as compared to equally spaced quantiles in Schefzik (2011).

3. *Arrange according to the multivariate structure:*
   In order to copy the multivariate structure of the ensemble, we determine the order statistics of the unprocessed ensemble members for each wind component, where ties are resolved at random. By this procedure, our ensemble becomes $U_{\sigma(1)}, \ldots, U_{\sigma(8)}$ and $V_{\mu(1)}, \ldots, V_{\mu(8)}$ with the respective order statistics $\sigma(1), \ldots, \sigma(8)$ and $\mu(1), \ldots, \mu(8)$. Then the accordingly ordered samples

$$\begin{pmatrix} \hat{U}_{\sigma(1)} \\ \hat{V}_{\mu(1)} \end{pmatrix}, \ldots, \begin{pmatrix} \hat{U}_{\sigma(8)} \\ \hat{V}_{\mu(8)} \end{pmatrix}$$

   constitute the ECC ensemble, which has inherited the dependence structure from the raw ensemble, although relying only on one-dimensional postprocessing.

In the general ECC approach, a sample of the magnitude of the ensemble size is generated and ordered; we follow this in the computation of the Euclidean error, the determinant sharpness and the multivariate rank histogram. For computing the energy score, however, a sample size of 10,000 is required (refer to Section 2.3.2). Thus we repeat the last two steps 1,250 times, thereby creating a total of 10,000 vectors.

5.3.5 Converting Ensemble and Density Forecasts

In order to compare all forecasts in this case study, we use the assessment tools introduced in Chapter 2. Some of the tools only apply to one form of probabilistic forecasts, either ensembles or probability distributions. Therefore, we have to convert the forecasts to assess their performance with these methods.

Sampling from a predictive distribution yields a discrete ensemble of any desired size. Gneiting et al. (2005) proposed taking quantiles at levels $\frac{i}{M+1}$, for $i = 1, \ldots, M$. However,

bivariate quantiles are hard to determine, which is why we use random samples to obtain the ensemble vectors.

For a given ensemble forecast, we replace it by a bivariate normal density estimate. The distribution characteristics are directly gained from the ensemble, with the ensemble means as mean vector, and the empirical ensemble variances and correlation generating the variance-covariance matrix of the new full predictive distribution.

5.4 Predictive Performance for Wind Vector Forecasts

At this point, we begin describing the predictive performance of the new bivariate EMOS method as well as of the other forecasting and postprocessing techniques introduced in the previous section. First, we recapture the example from Chapter 4, where we considered forecasts for the 23 January 2008 at Sea-Tac Airport. Then the performance of the local techniques at this station is discussed, by means of the forecasts for the calendar year 2008. Finally, we present our findings aggregated over all locations for both regional and local forecasts.

5.4.1 Results for Sea-Tac Airport, 23 January 2008

To illustrate EMOS in practice, we consider one date at one location, where we produce local and regional postprocessed forecasts. In Chapter 4, it was explained how we obtain the parameter estimates for the predictive distribution using linear regression for the mean vector, weighted non-linear regression for the correlation coefficient and maximum likelihood estimation for the variances. This estimation procedure was conducted for predictions at Sea-Tac Airport, valid on 23 January 2008.

Table 5.2 shows the parameter estimates for the local and the regional technique. For comparability reasons, we choose the regional EMOS forecast which uses simple linear regression on the ensemble mean. Correspondingly, the characteristics of the predictive distributions, as well as the ensemble means, ensemble variances and verifying observation are specified in Table 5.3.

Table 5.2: Parameter estimates at station KSEA, valid on 23 January 2008, for the local and regional EMOS techniques both using simple linear regression

Method	a_U	b_U	c_U	d_U	a_V	b_V	c_V	d_V
local EMOS	−0.69	0.86	2.29	2.67	−0.42	1.08	5.01	0.00
regional EMOS	−0.56	0.55	3.19	0.86	0.04	0.60	4.44	1.10

Table 5.3: Local and regional EMOS means, variances and correlation, as well as ensemble means, ensemble variances and the verifying observation for station KSEA, valid on 23 January 2008

Method	μ_U	μ_V	σ_U^2	σ_V^2	ρ
local EMOS	−0.25	−0.43	2.79	5.01	0.07
regional EMOS	−0.28	0.03	3.35	5.19	−0.29

$\overline{U}$	$\overline{V}$	S_U^2	S_V^2	U	V
0.51	−0.01	0.19	0.69	−1.34	0.77

Both methods roughly agree on the parameters for the predictive mean of the U component, while there is a significant difference in the other dimension. The local method suspects a large negative bias for V, and therefore corrects the mean in the wrong direction. In the case of the variance parameters, regional EMOS detects a relatively weak spread-skill relationship in both dimensions, with d_U and d_V being around one. Using only the location-specific training data even yielded that, for V, ensemble spread and forecast error are not correlated at all, resulting in a large value for c_V and a d_V equalling zero. In this way, the variances of the underdispersive ensemble are substantially enlarged. Also, by means of the correlation estimates, one can see the effects of using different predictive correlation functions for the two techniques.

Figures 5.6 and 5.7 let us compare the forecasts graphically. The regional EMOS mean is a better point forecast for the observation than the local EMOS mean. Their dispersion is very similar, apart from the differing correlations. For the local method, the observation lies just outside the 25% ellipse, while it lies well within for the regional method. Both plots show

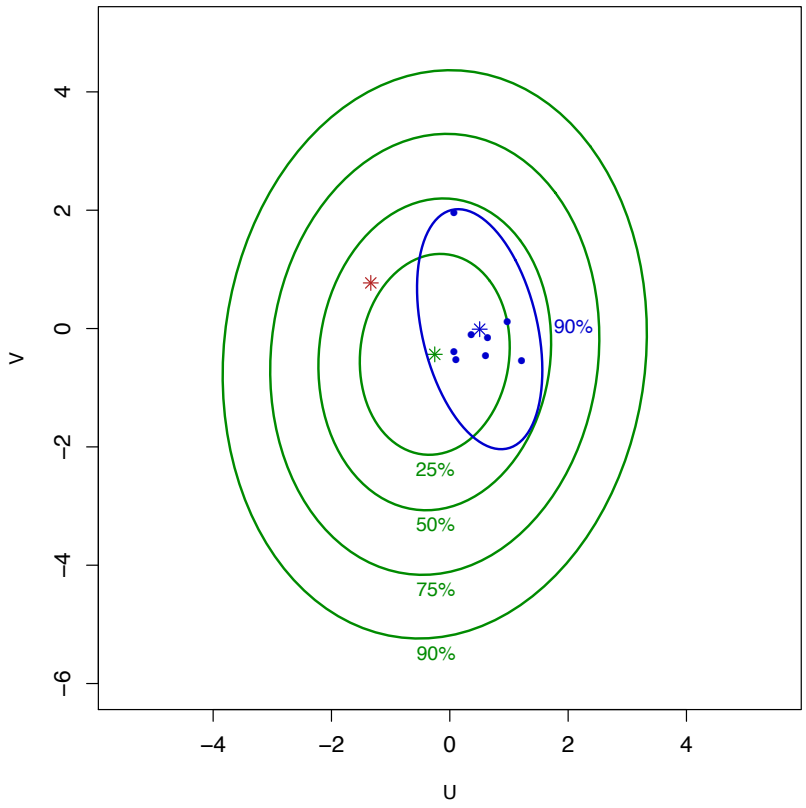


Figure 5.6: Contour plot of the local EMOS and the UWME forecasts; the green, blue and red asterisks represent the EMOS mean, the ensemble mean and the verifying observation, respectively; the green ellipses are the 25%, 50%, 75% and 90% EMOS prediction ellipses, while the blue one is the 90% prediction ellipse of the fitted UWME distribution

clearly that the spread of the UWME is far too small, with the observation way outside of the ensemble range and the 90% prediction ellipse.

When we visualise the predictive distributions with a perspective plot (Figures 5.8 and 5.9), the underdispersion becomes even more obvious, as the fitted UWME density is much more concentrated. Additionally, in Figure 5.8, the observation is positioned near the centre of the density, whereas in Figure 5.9 it lies in the outer margins, indicating the skill of the EMOS mean as point forecast.

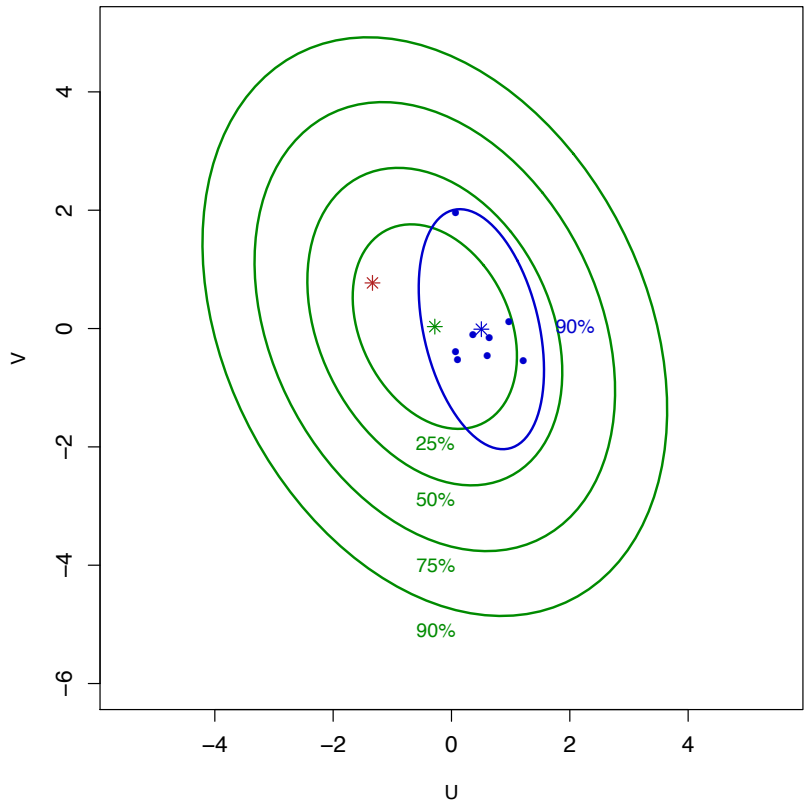


Figure 5.7: Contour plot of the regional EMOS and the UWME forecasts; the green, blue and red asterisks represent the EMOS mean, the ensemble mean and the verifying observation, respectively; the green ellipses are the 25%, 50%, 75% and 90% EMOS prediction ellipses, while the blue one is the 90% prediction ellipse of the fitted UWME distribution

Quantifying the performance using the energy score and the Euclidean error (Table 5.4), we find that, as expected, the regional method has the smallest scores, with local EMOS being slightly higher. The UWME is by far the least skillful predictor, in terms of both probabilistic and deterministic forecasts.

Table 5.4: Energy score and Euclidean error for local EMOS, regional EMOS, and UWME forecasts at station KSEA, valid on 23 January 2008

Method	ES	EE
Local EMOS	1.122	1.557
Regional EMOS	1.008	1.269
Raw ensemble	1.690	2.116

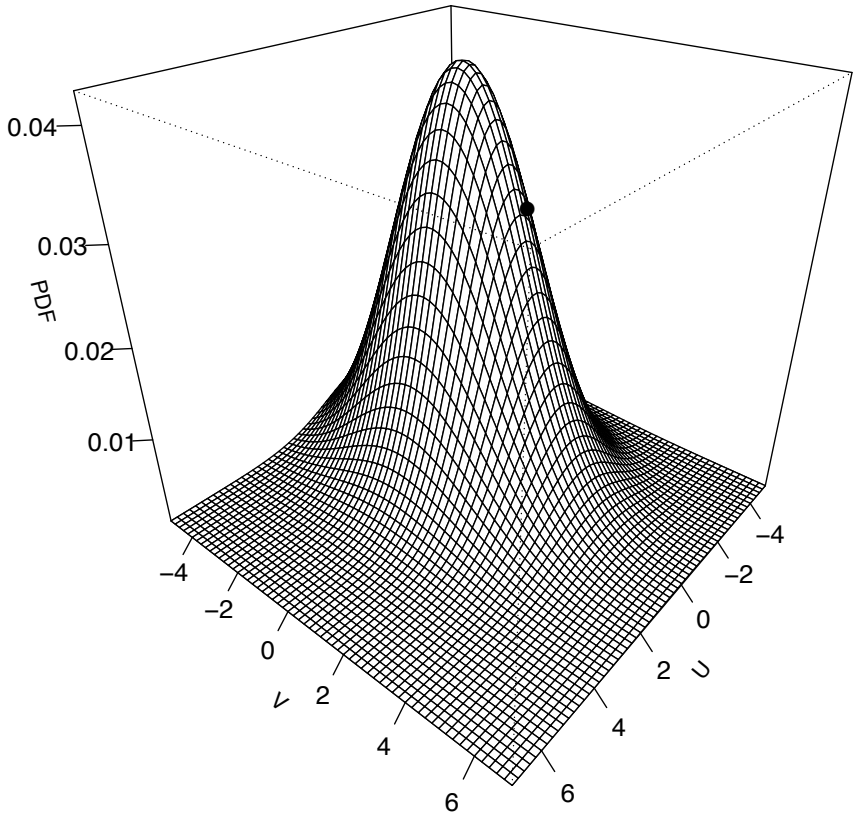


Figure 5.8: Three-dimensional perspective plot of the local EMOS predictive density; the dot indicates the verifying observation

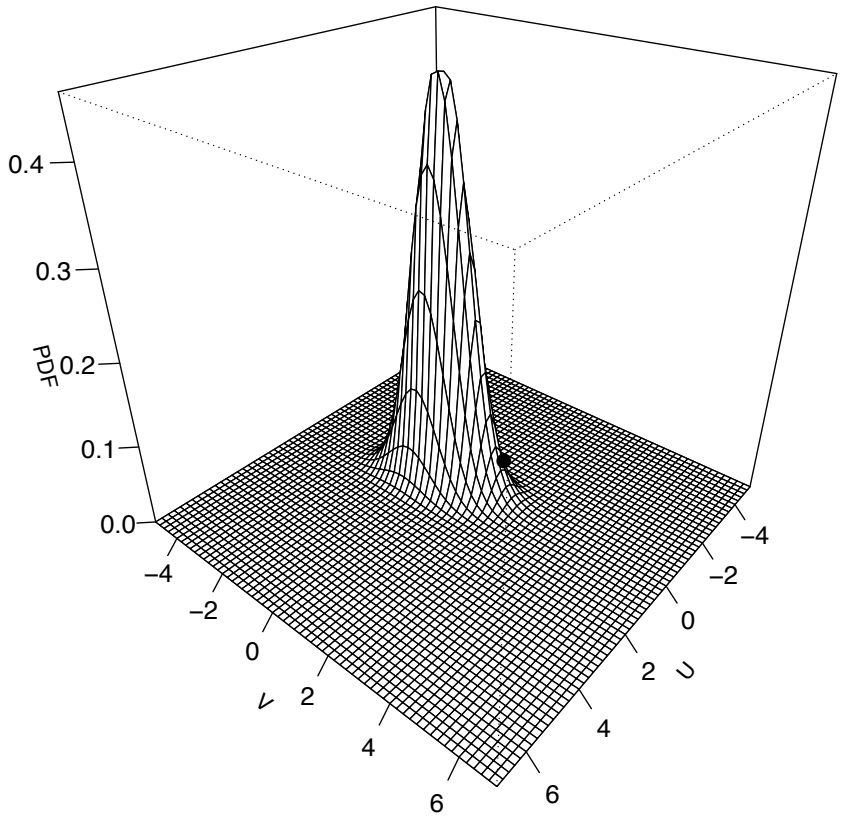


Figure 5.9: Three-dimensional perspective plot of the predictive density fitted to the ensemble forecast; the dot indicates the verifying observation

5.4.2 Results for Sea-Tac Airport, 1 January - 31 December 2008

In this section, we are interested in the performance of location-specific forecasts, as they can be crucial in order to predict local weather patterns that are not captured by the regional variant. Therefore, we want to compare our local technique to the local climatological forecast and a second postprocessing method, local error dressing. All forecasts use the same training period, that is the last 40 days where data was available. To ensure numerical stability when fitting normal distributions, the climatological and error dressing ensembles have an increased size of 40.

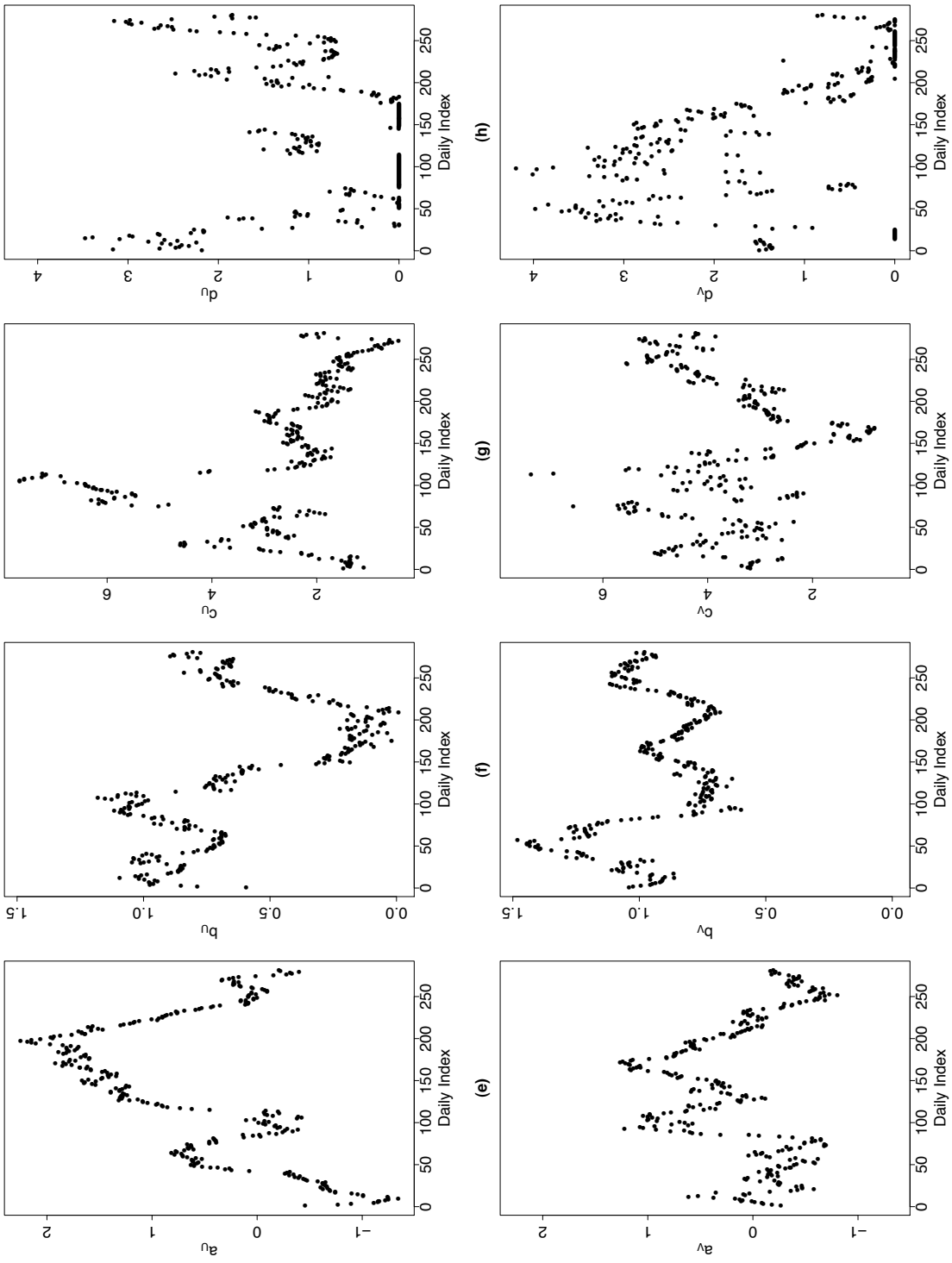


Figure 5.10: Parameter estimates for local EMOS forecasts at Sea-Tac Airport in 2008: (a) and (b) mean coefficients a_U and b_U, (c) and (d) variance coefficients c_U and d_U, (e) and (f) mean coefficients a_V and b_V, (g) and (h) variance coefficients c_V and d_V

Table 5.5: Comparison of the three local forecasting techniques and the raw ensemble in terms of proper scoring rules, calibration and sharpness, averaged over all dates in 2008 at Sea-Tac Airport

Method	ES	EE	Δ	DS
Local EMOS	1.944	2.743	0.069	1.990
Local climatology	2.553	3.610	0.028	2.574
Local error dressing	1.984	2.754	0.008	2.126
Raw ensemble	2.253	2.767	0.537	0.757

First we take a look at the estimated EMOS parameters for this station. In Figure 5.10, we see the development of each parameter over time, for both U and V components. Generally, the coefficients for V seem a little higher than the ones for U. The relationship between coefficients a_U and b_U, and between a_V and b_V is anti-cyclical, in that b becomes large when a becomes small and vice versa. Via the variance parameters, we can gather information about the spread-error correlation of the ensemble. For U, the parameter d decreases at first, while c increases. This implies a weakening relationship between ensemble spread and forecast error. Towards the end of the test period, c becomes very small, a sign for an improved ensemble skill. Conversely, for the V component, d is around zero for the last 50 days, whereas c has substantial values throughout the year. The somewhat high variance parameters reveal yet again that the ensemble spread is much too small and has to be enlarged.

Figure 5.11 shows the multivariate rank histograms for the three mentioned forecasts and the unprocessed ensemble, aggregated over all dates in 2008 at station KSEA. While the raw ensemble is underdispersive, indicated by the U-like shape, the other forecasts seem reasonably well calibrated, considering the fact that there are only 291 forecast cases. As a measure for calibration, the reliability index Δ in Table 5.5 quantifies and confirms the results of the histograms. Although the EMOS histogram looks more uniform than the one for the climatological forecast, it has actually a higher reliability index. This inconsistency originates in the fact that the size of the climatological (and the error dressing) ensemble is not an integral multiple of the UWME's size. So, by creating histograms with an equal number of

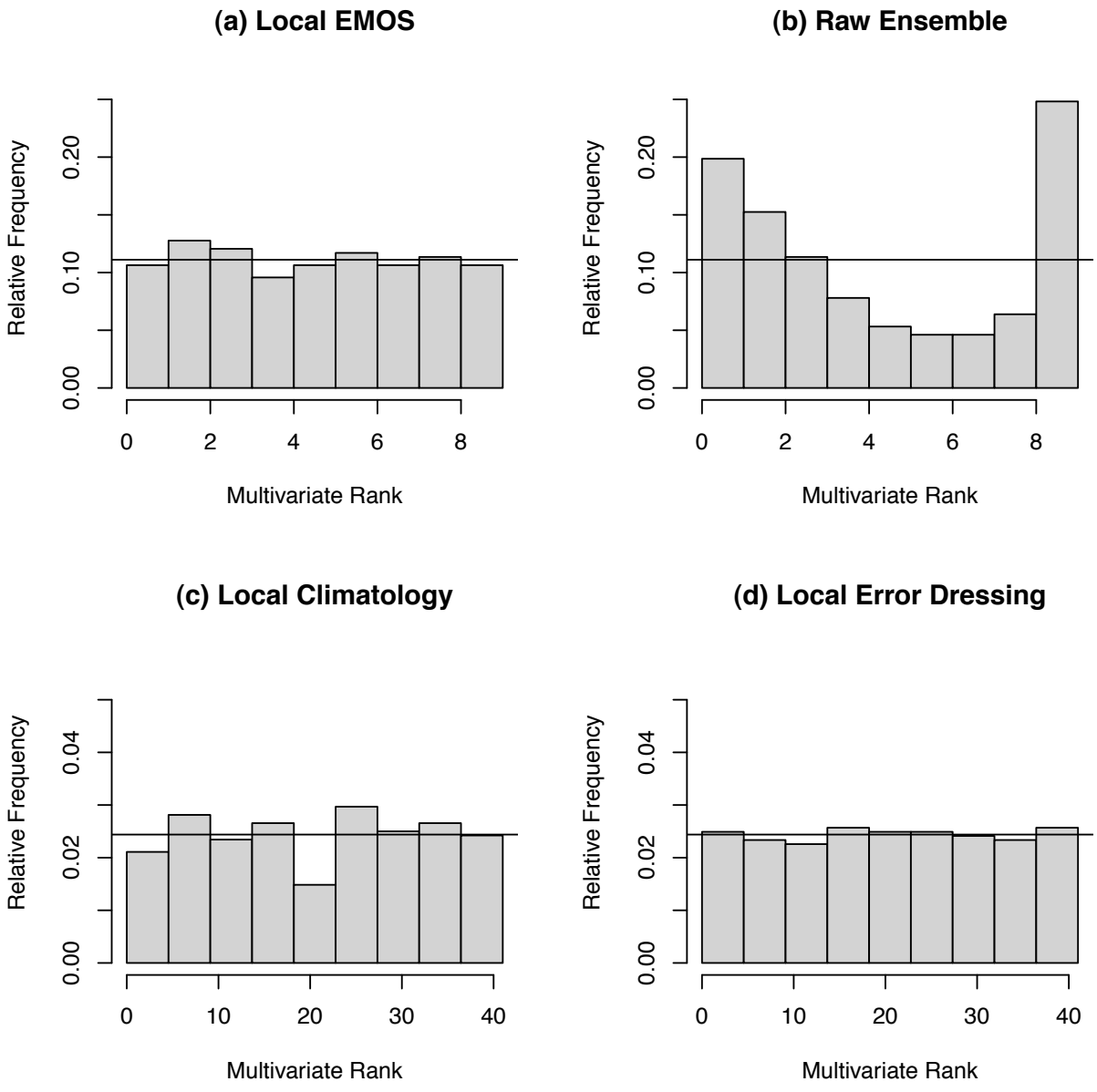


Figure 5.11: Multivariate rank histograms for location-specific forecasts at Sea-Tac Airport in 2008: (a) local EMOS forecast, (b) raw ensemble forecast, (c) local climatological forecast, and (d) local error dressing forecast

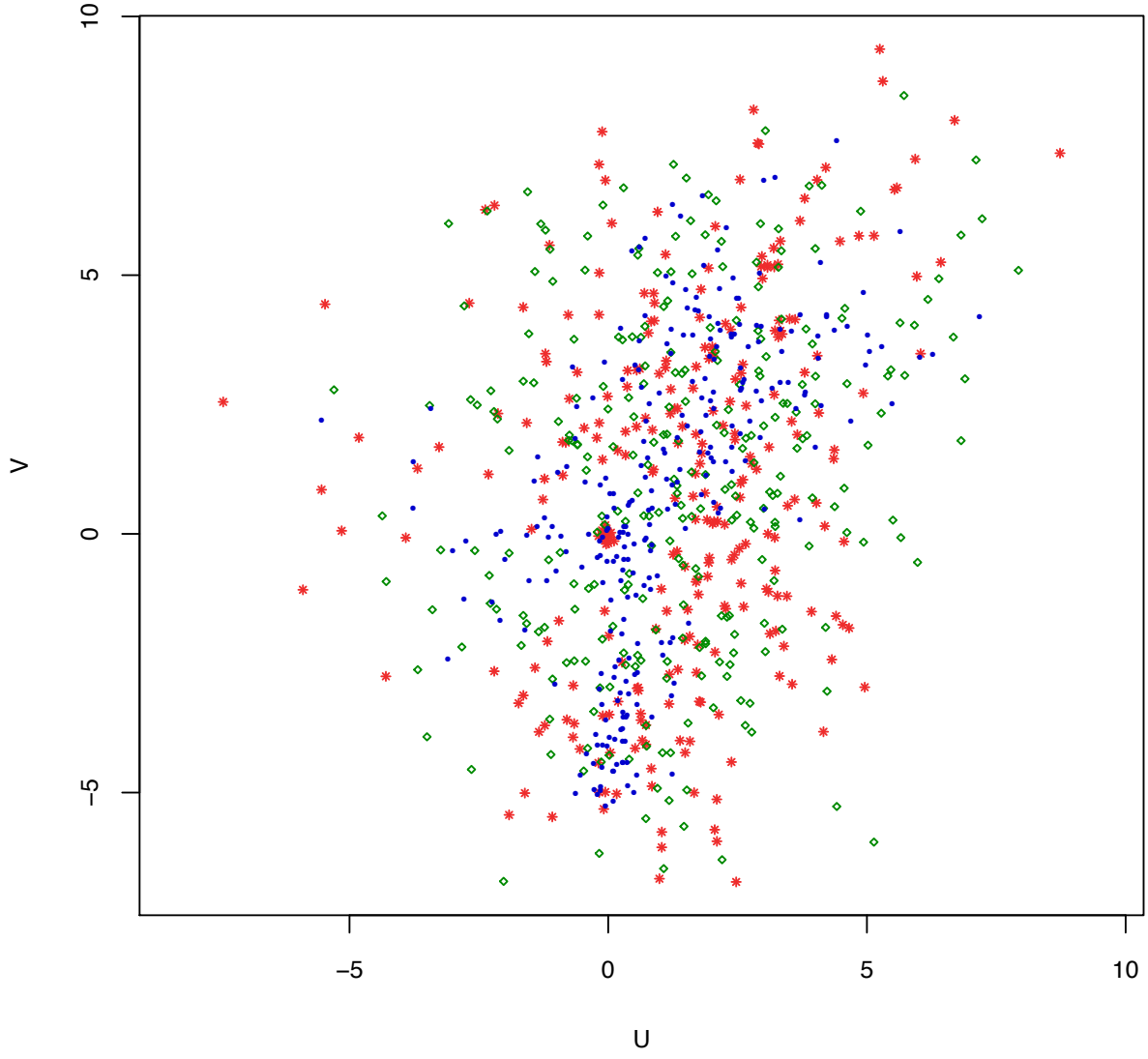


Figure 5.12: Marginal calibration diagram for Sea-Tac Airport; each red asterisks represents an observations, each blue dot a random ensemble member and each green diamonds a random vector from the EMOS predictive distribution

cells, the actual grade of calibration can become distorted. This applies to all climatology and error dressing histograms in the case study.

Local EMOS is, apart from the uncalibrated raw ensemble, the sharpest forecast, with the climatological ensemble the least sharp. The Euclidean error shows that the EMOS predictive median yields the best point forecast, being slightly better than the error dressing ensemble. Finally, in terms of the energy score, which addresses calibration and sharpness simultaneously, EMOS has the best overall predictive skill, closely followed by the local error dressing approach. The climatological forecast is not competitive, as it is too widely spread and has an even higher energy score than the raw ensemble.

In order to check EMOS for marginal calibration, we consider the marginal calibration diagram in Figure 5.12. As the verifying observations are discretised, we added a small random noise to each observed vector to simulate the undiscretised distribution and to improve interpretability. Comparing the empirical distributions, we see that, apart from some outliers, the local EMOS forecast reproduces the observation pattern. Although the UWME has no recognisable biases, it fails to predict the observations in the lower right and upper left and is too concentrated. So, it can be said that the local EMOS method is marginally calibrated for this station, while the spread of the ensemble is once more not large enough.

5.4.3 Results for the Pacific Northwest, 1 January - 31 December 2008

After examining the performance for one location, we now turn to the results aggregated over all stations in the Pacific Northwest. First we will address the issue of using only the ensemble mean or all ensemble members as predictors for the EMOS mean, as described in 5.2.1.1. Then our new method will successively be compared to the competing forecasts introduced earlier in this chapter.

5.4.3.1 Comparing Simple and Multiple Linear Regression

For the regional EMOS technique, we proposed to employ multiple linear regression based on the ensemble member forecasts in order to estimate the predictive mean. To ensure interpretable

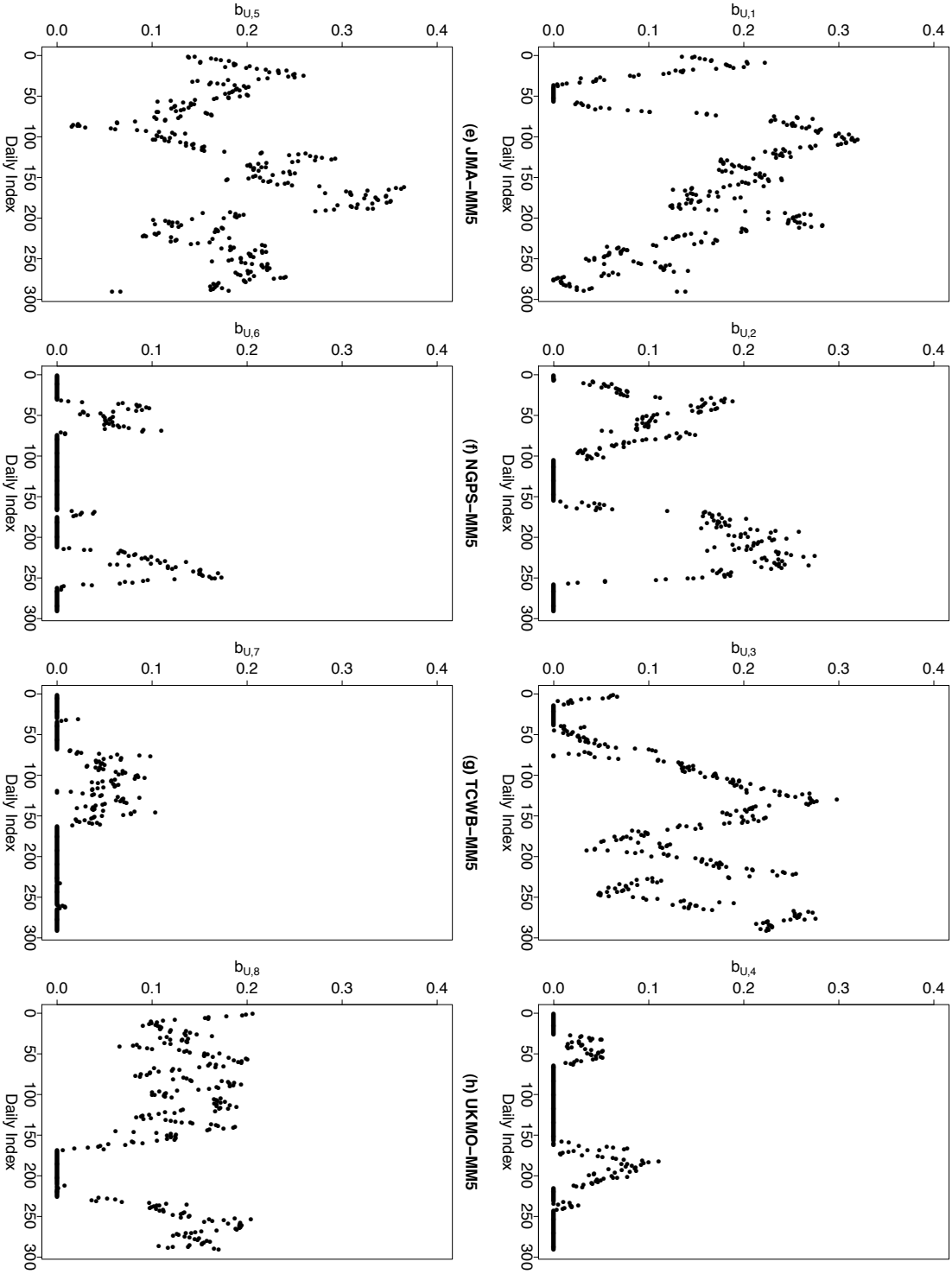


Figure 5.13: Parameter estimates for the regional EMOS$^+$ method using multiple linear regression, over the Pacific Northwest in 2008: (a)-(h) member model weights $b_{U,1}, \ldots, b_{U,8}$

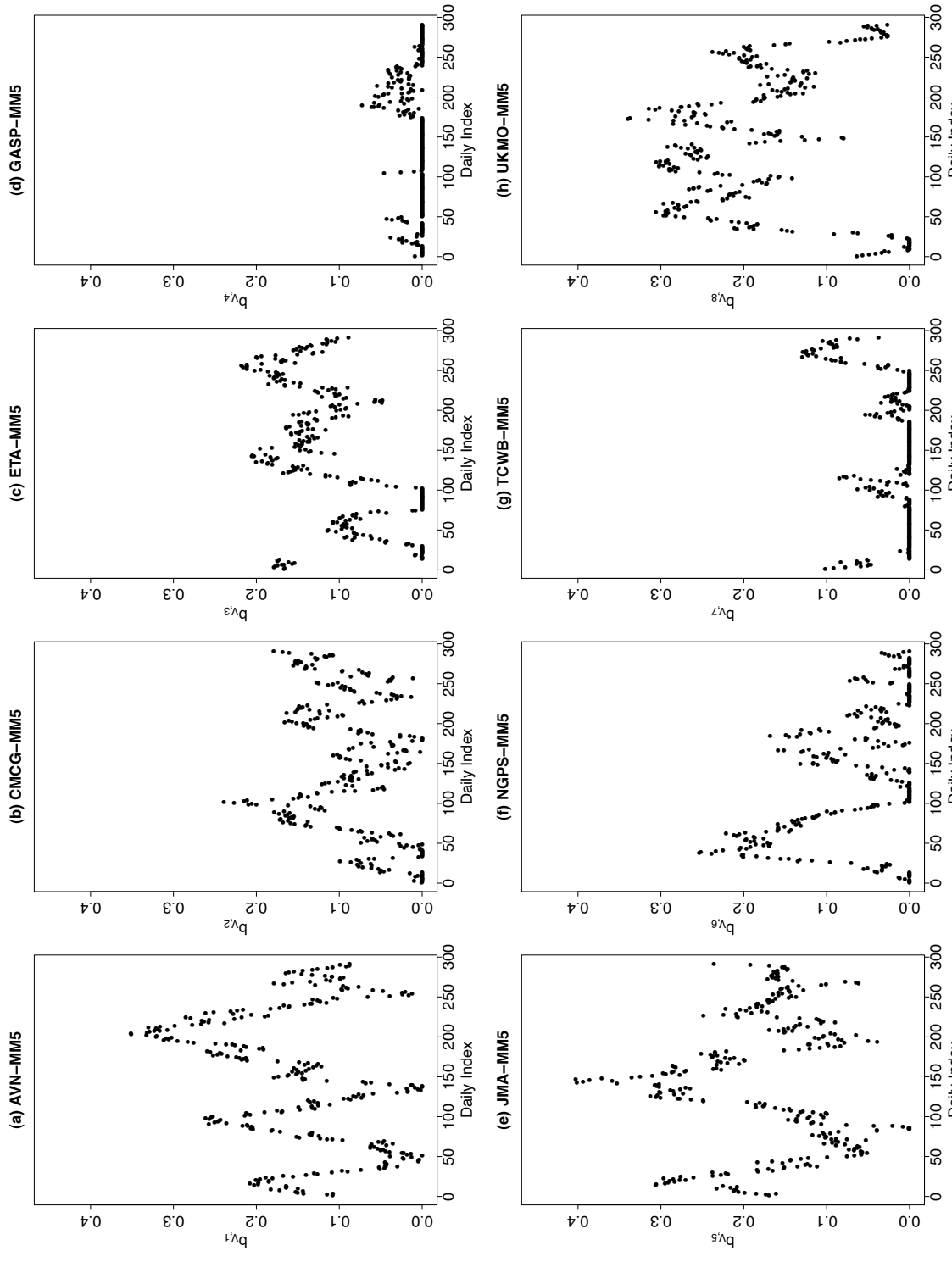


Figure 5.14: Parameter estimates for the regional EMOS$^+$ method using multiple linear regression, over the Pacific Northwest in 2008: (a)-(h) member model weights $b_{V,1}, \ldots, b_{V,8}$

member weights $b_{U,1}, \ldots, b_{U,8}$ and $b_{V,1}, \ldots, b_{V,8}$, we employ a variant of EMOS$^+$, and eliminate predictors from the regression equation until there are only non-negative coefficients left.

When we take a look at these coefficients in Figures 5.13 and 5.14, the predictive skill of the individual member models can directly be compared. For both components, the AVN-MM5 and JMA-MM5 forecasts received the highest weights, which corresponds well with the results in Thorarinsdottir and Gneiting (2010), obtained over the same test period. The forecasts generated using the initial conditions by the Australian Bureau of Meteorology, GASP, and by the Taiwan Central Weather Bureau, TCWB, had the lowest weights throughout the year, being zero most of the time. Interestingly, the NGPS-MM5 member shows almost no skill for the U component, but gets rather solid weights for V. Although the AVN-MM5 and the ETA-MM5 forecasts are obtained with initial data from the same source, the National Centers for Environmental Prediction, their coefficient estimates are both substantial over the whole test period, which indicates that they are not highly correlated.

In Figures 5.15 and 5.16, we see the estimated parameters for the intercept and the predictive variances. We notice that, in comparison to the local parameters in Figure 5.10, d_U and d_V never vanish. This increased stability is due to the higher amount of training data available for the regional technique. The ensemble for the V component shows a much more prominent spread-error correlation than the one for the U component, as the values for c_V are very small, while d_V tends to be large. Again, the coefficients c_U and d_U exhibit an anti-cyclical behaviour, suggesting a varying relationship between ensemble spread and forecast skill.

Considering the scores in Table 5.6, for both simple and multiple regression techniques, and the multivariate rank histograms in Figure 5.17, we notice a small improvement in performance when using all ensemble members to predict the EMOS mean. Not only implies the lower Euclidean error a more accurate deterministic forecast, but the EMOS$^+$ version is also better calibrated and sharper, leading to a lower energy score. The results show that the assumption of exchangeable ensemble members does not apply to the UWME, although it apparently causes only a minor loss in predictive skill to use the ensemble mean as sole predictor. Therefore, it seems worthwhile to keep simple linear regression for the local approach, where we have less training data, and adopt multiple linear regression for the regional technique.

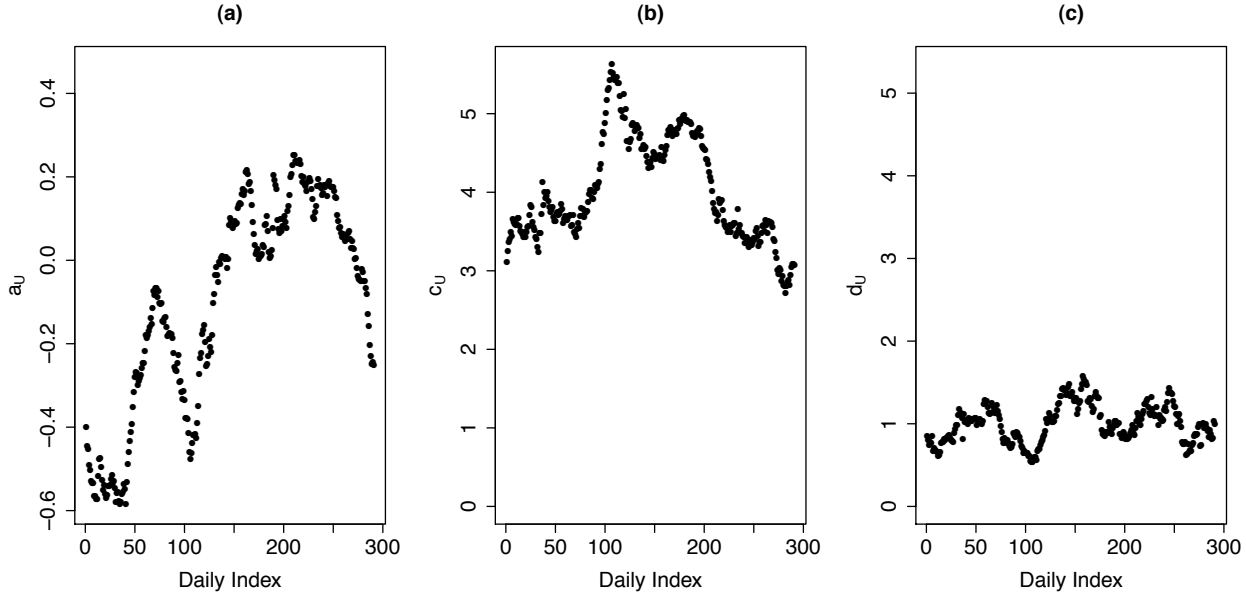


Figure 5.15: Parameter estimates for the regional EMOS$^+$ method using multiple linear regression, over the Pacific Northwest in 2008: (a) intercept a_U, (b) and (c) variance coefficients c_U and d_U

5.4.3.2 Results for Regional Forecasts

In this section, we assess the performance of the regional EMOS$^+$ method as compared to other regional forecasting techniques and the unprocessed UWME ensemble. We concentrate on the multiple regression version and give results in terms of the energy score, Euclidean error, multivariate rank histogram and determinant sharpness.

Looking at the histograms in Figures 5.17 and 5.18, we find that the bivariate EMOS$^+$, climatological and error dressing forecasts, as well as the ECC ensemble, generated using the regional EMOS^{UV} distributions, are nearly calibrated, while the regional EMOS^{UV} method itself is overdispersive. The latter is however not surprising, as this technique predicts independent vector components and the respective variances are estimated individually. Therefore, in order to capture outlying observations, the spread is enlarged in both dimensions, and by that overestimated. The multivariate histogram for the UWME in Figure 5.20 shows the raw ensemble's underdispersion clearly, as seen before when we considered only one station.

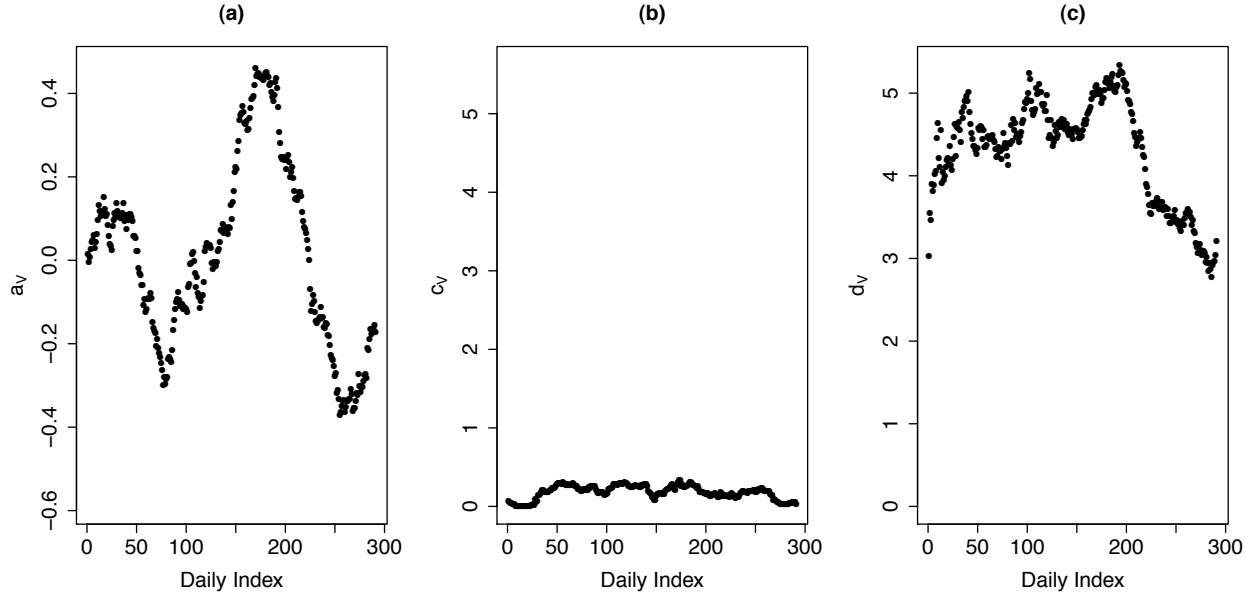


Figure 5.16: Parameter estimates for the regional EMOS$^+$ method using multiple linear regression, over the Pacific Northwest in 2008: (a) intercept a_V, (b) and (c) variance coefficients c_V and d_V

Our findings are confirmed by looking at Table 5.6. The calibrated forecasts have a similar reliability index and can be ranked by their sharpness, according to the principle of maximising the sharpness subject to calibration. EMOS^{UV} has, as surmised, the by far highest determinant sharpness value, while the UWME ensemble has the lowest of all. However, these forecasts are not calibrated and their overall performance in terms of the energy score is not competitive. Of the calibrated methods, ECC is the sharpest, followed by EMOS$^+$ and the error dressing ensemble, while the climatological forecast is much more spread out.

When we look at the skill of the respective bivariate medians as deterministic forecasts, EMOS$^+$ has the lowest Euclidean error. Although EMOS^{UV} is uncalibrated, it produces a very good point forecast, almost equalling the best score. ECC and error dressing perform about similarly well, but the climatological ensemble is even worse than the raw ensemble, which might explain the very high energy score. Apart from this issue, the uncalibrated UWME and

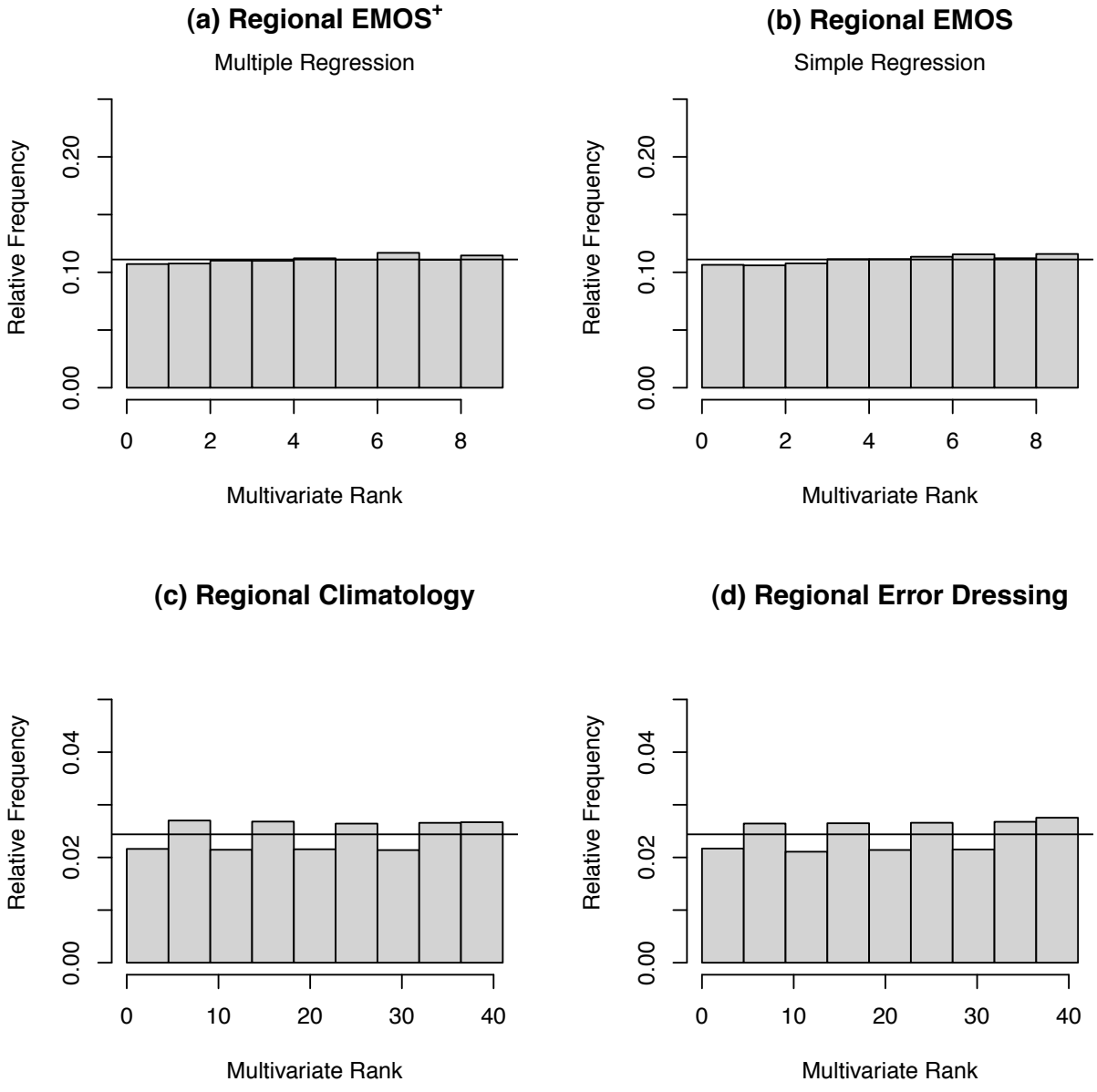


Figure 5.17: Multivariate rank histograms for regional forecasts over the Pacific Northwest in 2008: (a) regional EMOS$^+$ forecast using multiple linear regression, (b) regional EMOS forecast using simple linear regression, (c) regional climatological forecast, and (d) regional error dressing forecast

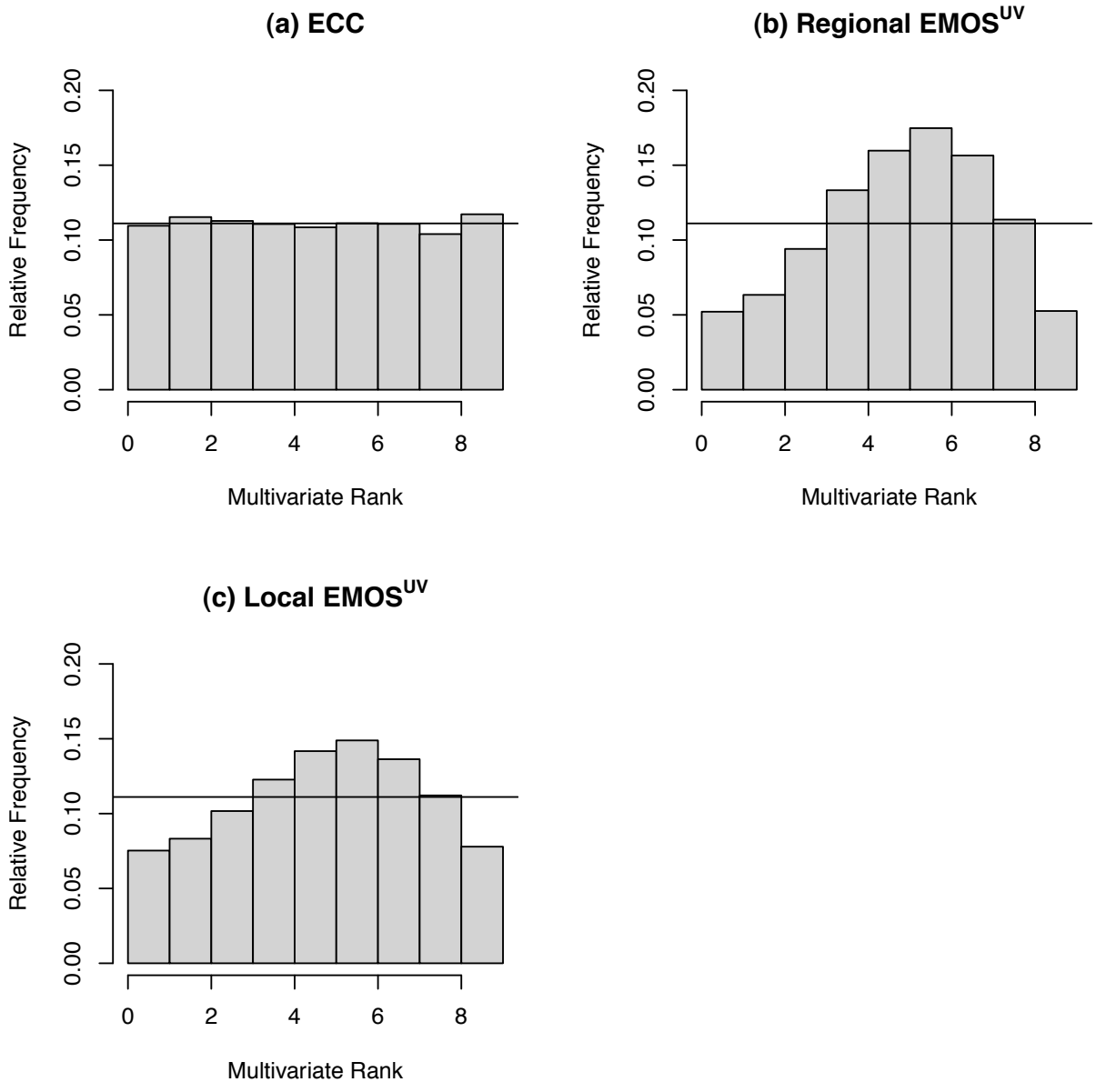


Figure 5.18: Multivariate rank histograms for competing forecasts over the Pacific Northwest in 2008: (a) ECC ensemble forecast based on regional EMOS^{UV}, (b) regional EMOS^{UV} forecast, and (c) local EMOS^{UV} forecast

Table 5.6: Comparison of the regional forecasting techniques and the raw ensemble in terms of proper scoring rules, calibration and sharpness, averaged over the Pacific Northwest in 2008

Method	ES	EE	Δ	DS
Regional EMOS$^+$	2.005	2.792	0.020	2.254
Regional EMOS	2.011	2.802	0.026	2.272
Regional climatology	2.677	3.632	0.019	2.924
Regional error dressing	2.181	3.002	0.023	2.380
ECC ensemble	2.325	2.998	0.024	1.859
Regional EMOS^{UV}	2.432	2.798	0.365	4.701
Raw ensemble	2.467	3.006	0.528	0.819

EMOS^{UV} expectedly receive the largest energy score values. The bivariate EMOS$^+$ method clearly performs better than the other techniques.

Surprisingly, the ECC ensemble has a relatively high energy score, though being calibrated and the sharpest forecast of all. In Section 2.2, we mentioned that the determinant sharpness tends to overly reward strongly calibrated forecasts, although the variances might be large. Figure 5.19 pictures the predictive correlation of the regional EMOS$^+$ method and the empirical correlation for the ECC ensemble at one station. This graphical comparison shows that ECC predicts correlations ranging from -1 to 1, while the EMOS$^+$ correlation always lies between -0.4 and 0.1. We can conclude that these high correlations estimates distort the actual spread of the ECC approach and the true performance relative to EMOS$^+$ reveals itself in the energy score. Possibly, this effect is caused by the relatively small ensemble size of eight; in order to compute the determinant sharpness, we fit a normal distribution to the ensembles using the empirical variances and correlation. However, when applied to a large ensemble, the empirical correlation can be estimated more accurately.

Summing up, we notice that the two-dimensional EMOS$^+$ method clearly outperforms the other regional forecasting and postprocessing techniques. EMOS^{UV} overestimates the bivariate spread, and when additionally combined with the ECC approach, overestimates the correlation. Our way of predicting the correlation depending on the predicted wind direction

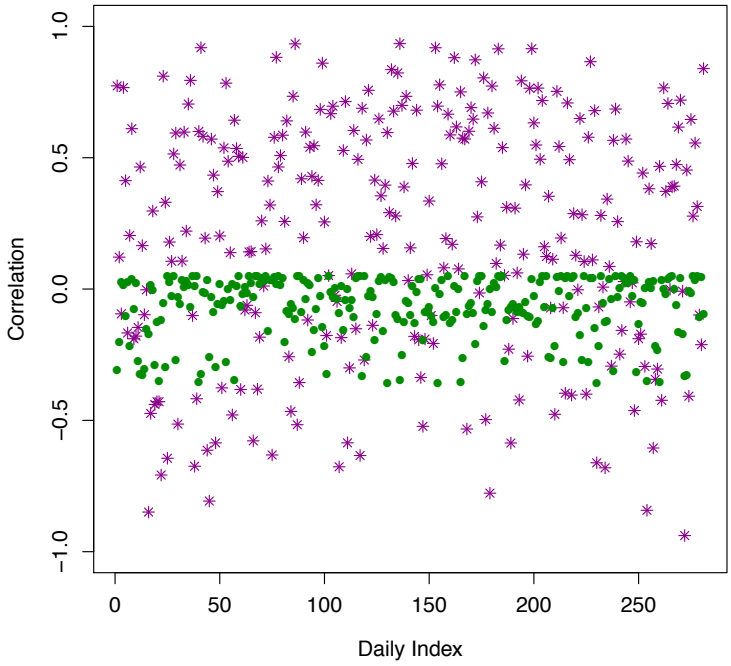


Figure 5.19: Predicted correlation at Sea-Tac Airport in 2008, the green dots represent the regional EMOS$^+$ and the magenta asterisks the ECC method

seems to be essential for the good performance of EMOS$^+$. Though the climatological forecast is calibrated, its median is a poor deterministic forecast and on average, the prediction ellipses are too wide. The error dressing method produces good results, but is still not as sharp as EMOS$^+$ and therefore has a higher energy score.

5.4.3.3 Results for Local Forecasts

After having seen results for the predictive skill of the local EMOS technique at Sea-Tac Airport, we now want to assess the performance for all 79 stations in the Pacific Northwest. In Figures 5.18 and 5.20, the multivariate histograms for the different local approaches are provided. Although the local EMOS^{UV} technique seems a little more calibrated than the regional version, it is still overdispersive, due to the lack of correlation between the vector

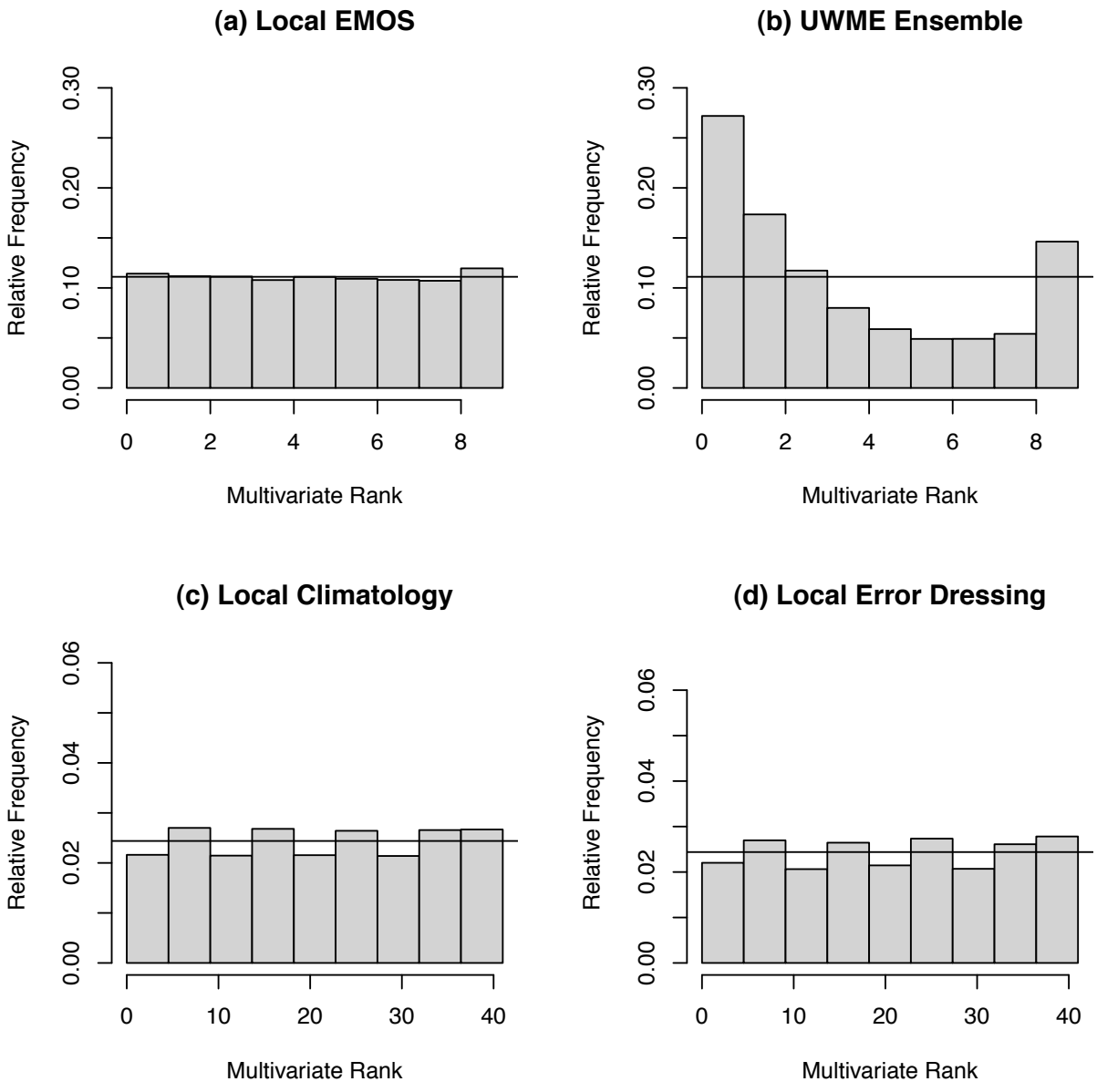


Figure 5.20: Multivariate rank histograms for different forecasts over the Pacific Northwest in 2008: (a) local EMOS forecast, (b) UWME forecast, (c) local climatological forecast, and (d) local error dressing forecast

Table 5.7: Comparison of the local forecasting techniques and the UWME in terms of proper scoring rules, calibration and sharpness, averaged over the Pacific Northwest in 2008

Method	ES	EE	Δ	DS
Local EMOS	1.875	2.608	0.025	1.906
Local climatology	2.404	3.361	0.023	2.421
Local error dressing	2.064	2.863	0.025	2.185
Local EMOS^{UV}	2.279	2.612	0.213	3.716
Raw ensemble	2.467	3.006	0.528	0.819

components. Again, EMOS, the climatological ensemble and the error dressing ensemble are calibrated, as can also be seen by means of the reliability index given in Table 5.7. We have established before that the UWME forecast is underdispersive, i.e uncalibrated, and even though its determinant sharpness is minimal, this results in a high energy score. Being overdispersed, EMOS^{UV} also is the least sharp forecast and not competitive in terms of the energy score, although its median has a very high deterministic forecast skill.

Only considering the calibrated methods, EMOS is sharper than error dressing and both are sharper than the climatological forecast. Therefore, EMOS also has by far the lowest energy score, followed by error dressing and climatology. As could be expected, all local methods perform better than the regional ones, especially in terms of sharpness. Looking at the Euclidean error, we find that the climatological forecast again seems to have no deterministic forecast skill, being even worse than the raw ensemble. The error dressing ensemble is somewhat worse than EMOS, which again produces the best point forecasts.

As with the regional method, we can conclude that EMOS performs better than all other forecasting and postprocessing methods, in terms of the energy score, the Euclidean error and the determinant sharpness. The localised approach resulted in an energy score improvement of 6.5% and an increase in sharpness of 15%. Though the climatological and error dressing ensembles are calibrated, they are not as sharp as EMOS and therefore have higher energy scores. EMOS^{UV} is again overdispersed and has too wide prediction ellipses, but shows some predictive skill as deterministic forecast.

Table 5.8: Mean CRPS and MAE for the different EMOS methods, averaged over the Pacific Northwest in 2008

Method	CRPS	MAE
Regional EMOS, univariate	1.172	1.638
Regional EMOS, bivariate	1.219	1.692
Local EMOS, univariate	1.133	1.564
Regional EMOS bivariate	1.134	1.577

Figures 5.21 and 5.22 provide a graphical comparison between regional and local forecasts at different locations. Generally, the difference between the spread of the U and V components seems to be larger in the local technique, as the ellipses are more stretched. Also, regional EMOS tends to predict negative correlation more often than local EMOS. The wind direction of the EMOS mean does not differ much for the two techniques and the performance at individual stations is very similar. Furthermore, we notice that the prediction ellipses of the local technique are on average smaller than those of the regional technique, confirming that the former is sharper than the latter.

5.5 Predictive Performance for Wind Speed Forecasts

The new bivariate EMOS method not only produces calibrated and sharp forecasts for wind vectors, it also provides a way to predict wind speed through the length of the vectors. In this section, we show how we can generate wind speed forecasts from bivariate predictive distributions and compare their performance to the univariate EMOS method for wind speed described in Section 3.3 and Thorarinsdottir and Gneiting (2010). As our wind vectors represent instantaneous wind, we have to apply the univariate method to instantaneous wind speed, as compared to maximum wind speed in the article. Also, we will give our results in $^{\mathrm{m}}/_{\mathrm{s}}$ rather than in knots. To ensure comparability, we restrict the analysis to dates and locations where both wind speed and wind vector forecasts and observations were available.

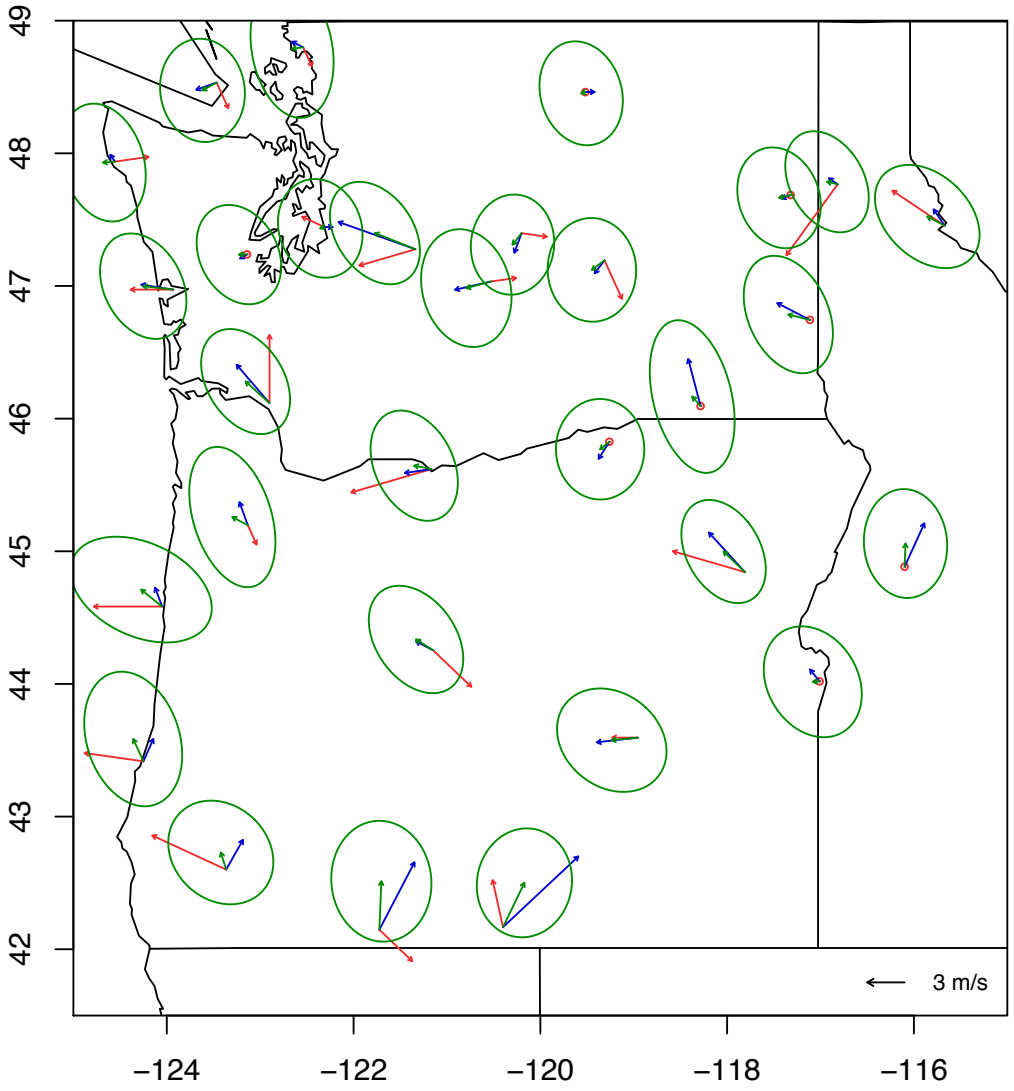


Figure 5.21: Map of the US states of Washington, Oregon and Idaho, showing the UWME means in blue, the regional EMOS means and 80% prediction ellipses in green, and the verifying observations in red, for a selection of 30 observational locations, valid on 23 January 2008

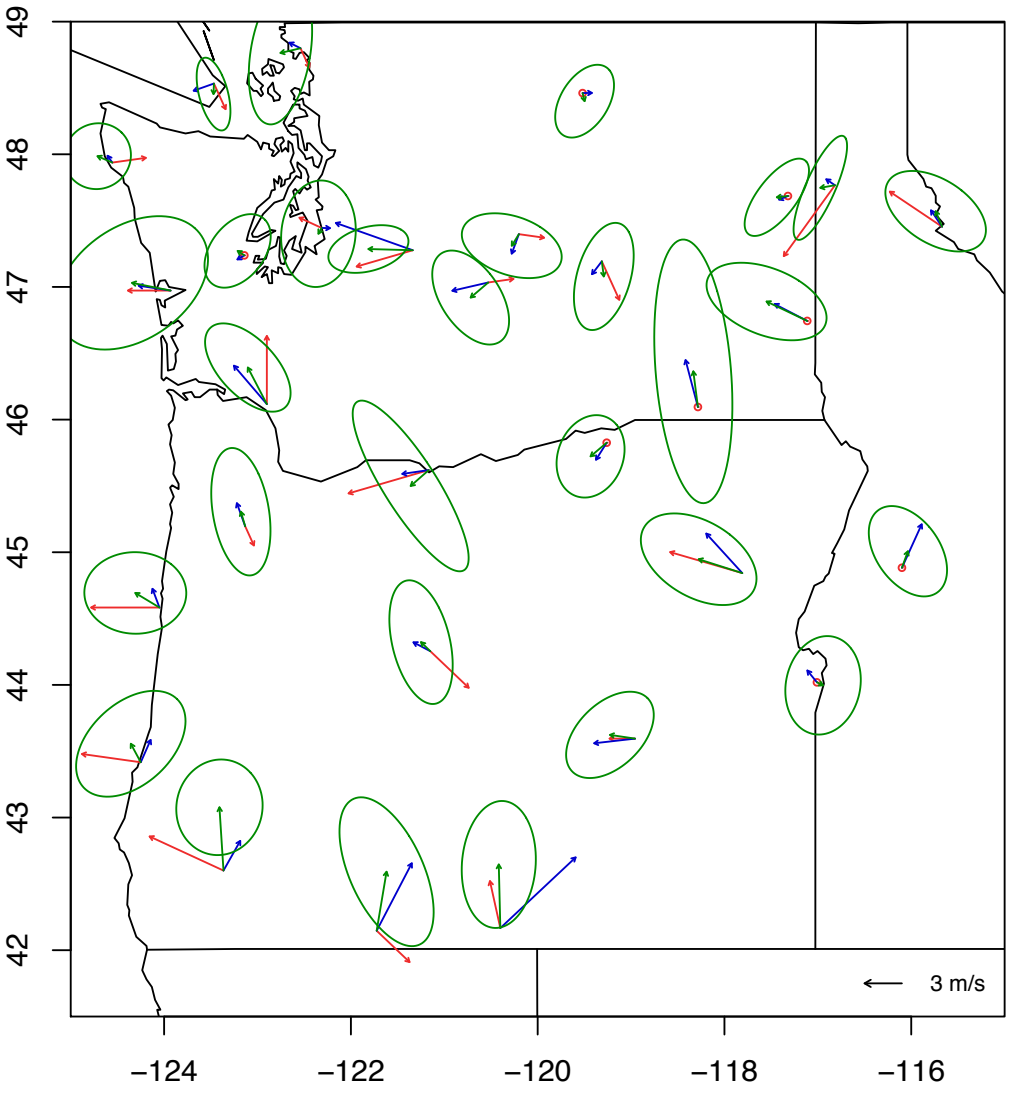


Figure 5.22: Map of the US states of Washington, Oregon and Idaho, showing the UWME means in blue, the local EMOS means and 80% prediction ellipses in green, and the verifying observations in red, for a selection of 30 observational locations, valid on 23 January 2008

First, we compute the univariate EMOS predictive distributions for the regional and the local techniques, following the suggestion of 20 and 40 days training period length, respectively. Then, in order to convert wind vectors into wind speed, we sample from the bivariate predictive distributions 100 times and apply the Euclidean norm to each vector, obtaining the vector length and therefore an ensemble of wind speed forecasts.

For assessing the predictive skill, the mean CRPS, mean MAE, PIT histogram and verification rank histogram are employed, in the way they are described in Chapter 2. In Figure 5.23, we see the four histograms for the individual techniques. At first sight, they don't seem to be calibrated and rather show a large spike at zero, which means that both approaches fail to predict very low wind speed. Probably this is caused by the discretisation of the wind speed observations (Section 5.1), especially as values smaller than 1 knot are recorded as 0. For maximum wind speed, this was only the case for 30 observations, whereas for instantaneous wind speed there are more than 2,700 observations equal to 0, i.e. in 14% of all forecast cases.

However, this problem only arises in the PIT and rank histograms. When we look at the results of the proper scoring rules, we see that, in terms of the CRPS, both univariate and bivariate methods perform equally well for the local technique, while the univariate EMOS median is a better deterministic forecast than the median of the ensemble created from the bivariate EMOS distribution. The difference is quite large for the regional techniques, where the univariate method outperforms the bivariate, in terms of both the CRPS and the MAE.

We can summarise that the two EMOS methods seem to have a certain bias, although this impression is probably due to discretisation of the observational data. The univariate technique has the higher predictive skill for probabilistic and point forecasts, although they perform equally well in the localised case. This is only natural, as it was specifically designed and optimised for wind speed, while the bivariate approach issues wind speed forecasts only as a byproduct.

Concluding this chapter, we note that the new bivariate EMOS method is a very efficient and successful way to postprocess ensemble forecasts for wind vectors. It corrects the underdispersion of the raw ensemble in order to produce calibrated bivariate distributions. An

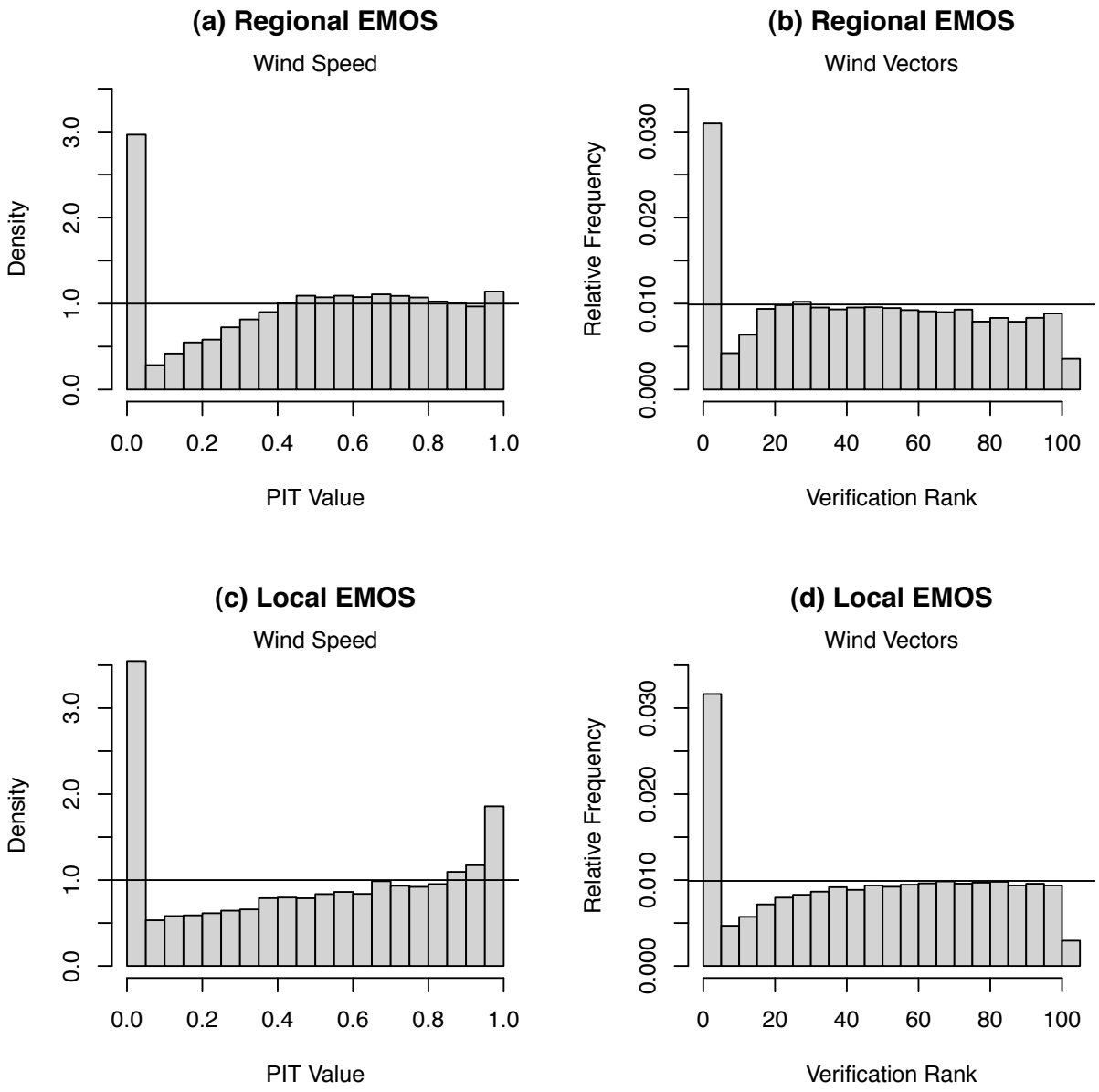


Figure 5.23: Calibration assessment for wind speed forecasts over the Pacific Northwest in 2008: (a) PIT histogram for the univariate regional EMOS method, (b) verification rank histogram for the bivariate regional EMOS$^+$ method, (c) PIT histogram for the univariate local EMOS method, and (d) verification rank histogram for the bivariate local EMOS method

important feature is the estimation of the correlation factor using the predicted wind direction, which results in a low determinant sharpness. In terms of proper scoring rules, EMOS received both the lowest energy score and Euclidean error. Although it might be slightly uncalibrated, the wind speed forecasts created by applying the Euclidean norm are almost as skillful as for the specialist univariate EMOS method.

Chapter 6

Summary and Discussion

In the introduction, we established that there is a huge need for probabilistic forecasts in the form of calibrated and sharp full predictive distributions. They can be obtained by statistical postprocessing of ensemble forecasts, which are generated by numerical weather prediction models and represent possible future states of the atmosphere. Most ensembles are underdispersive and don't account for all uncertainties involved in numerical weather prediction. Postprocessing techniques like EMOS or BMA yield very good results, as they improve the predictive performance of the unprocessed ensemble substantially.

So far, these methods were applied to univariate weather variables like temperature, air pressure or precipitation. When considering surface wind, one has only the possibility to address one-dimensional features like wind speed, wind direction or one of the zonal and meridional wind components. However, by employing univariate postprocessing, essential characteristics may not be captured. Therefore we proposed a bivariate variant of EMOS, which applies to the wind components jointly rather than separately and thus creates two-dimensional wind vector forecasts.

Empirical analysis of wind component data yielded that the observations, conditional on the ensemble forecasts, can be modelled with a bivariate normal distribution, thereby allowing for a similar approach as with univariate EMOS. A straightforward way to express the predictive mean vector in terms of the ensemble is to write the individual means as linear regression equations, with the ensemble mean or the ensemble member forecasts as predictors. The parameters are estimated on a sliding training period, typically of 30 or 40 days. In

order to make use of a possible spread-skill relationship, we employ an affine function of the ensemble spreads for the predictive variances. Here, maximum likelihood estimation ensures that calibration and sharpness of the predictive distribution are addressed simultaneously.

The main novelty of our approach lies in the prediction of the spatial correlation between the U and V wind components. During the above mentioned data analysis, we noticed that the empirical correlation of the observation vectors relates strongly to the predicted wind direction. Further examination yielded that this relationship approximately takes the form of a trigonometric curve, which we estimate with historic data, using weighted non-linear regression. In order to determine the predictive correlation for each forecast, this function is evaluated at the wind direction predicted by the bias-corrected mean.

We tested the performance of the new EMOS method for the University of Washington Mesoscale Ensemble, creating 48-h ahead forecasts over the Pacific Northwest. The resulting predictive distributions turned out to be calibrated and sharp, for both a regional and a local technique. We found that EMOS corrects the biases and the underdispersion of the ensemble, which results in a significantly improved forecast skill. Moreover, when compared to other well-calibrated forecasting techniques like the climatological and error dressing ensembles, EMOS proves to be the sharpest forecast and to have the lowest energy scores, indicating the best overall performance. Generally, the local technique is sharper than the regional technique and, as both are calibrated, receives the smaller energy score.

In order to assess the benefits of estimating the correlation coefficient by means of the wind direction, we applied univariate standard EMOS to the separate wind vector components and combined the results to a bivariate distribution under the assumption of independence. This approach however yields overdispersed forecasts, both in the regional and the local context. To include information gained from the raw ensemble, the ECC approach was employed. It performs best in terms of the determinant sharpness, but still EMOS has a substantially lower energy score. After examining the predicted and fitted correlation coefficients of EMOS and ECC, we noticed that this inconsistency is caused by the fact that the determinant sharpness overly rewards highly correlated forecasts. For ensembles with more members and therefore less statistical variability in the fitting of the variance-covariance matrix, ECC might produce better results.

For all techniques, the respective bivariate median was computed and its skill as a point forecast was assessed. Again, two-dimensional EMOS had the lowest Euclidean error, although the univariate EMOS^{UV} performed equally well.

As wind vector forecasts can easily be transformed into wind speed forecasts using the Euclidean norm, we tested the bivariate EMOS method against the specialist univariate method for instantaneous wind speed. Surprisingly, both techniques seem to be uncalibrated and fail to predict very low wind speed, which can be traced back to the discretisation of wind speed observations. In terms of the CRPS, the regional univariate method shows a better predictive performance than the bivariate one, but both local techniques are equally skillful. Considering the MAE, however, it becomes obvious that the univariate approach yields the best deterministic forecasts.

We conclude that the statistical postprocessing of wind vectors works best in a bivariate framework, where the spatial relationship between the components is taken into account. Our method models the correlation as a function of the predicted wind direction, which turned out to yield very good results, especially when compared to other techniques using only the information contained in the ensemble forecasts.

Previous work on the topic of wind vectors was done by Sloughter (2009), developing a two-dimensional version of Bayesian model averaging. Also, Pinson (2011) recently proposed a way to adaptively calibrate bivariate ensemble forecasts, which is in some points related to the EMOS method. The author likewise assumes wind vector observations to be bivariate normally distributed; however, instead of fitting full predictive distributions, the ensemble forecasts are corrected using translation and dilation factors only. These factors are obtained from model parameters similar to the coefficients for the EMOS predictive mean vector and predictive variances. For estimating the parameters, a recursive maximum likelihood approach with exponential forgetting of past observations is proposed, as compared to the conventional maximum likelihood estimation in the context of the EMOS method. The two methods also differ in the treatment of the correlation between the wind components; in Pinson (2011), a possible correction of the correlation is not considered, which works well for the here employed ECMWF ensemble. This ensemble consists of 50 members and surely provides more information about the spatial structure of wind vectors than the 8 member UWME. Therefore it would be interesting to compare both methods when using the same data set.

Finally, we want to take a look at possible improvements or extensions to the bivariate EMOS technique. In order to optimise the parameter estimation, an exponential forgetting method like in Pinson (2011) could be introduced, where the model parameters are successively updated with the newest set of forecasts and observations. This approach might result in a better ability to react to seasonal changes.

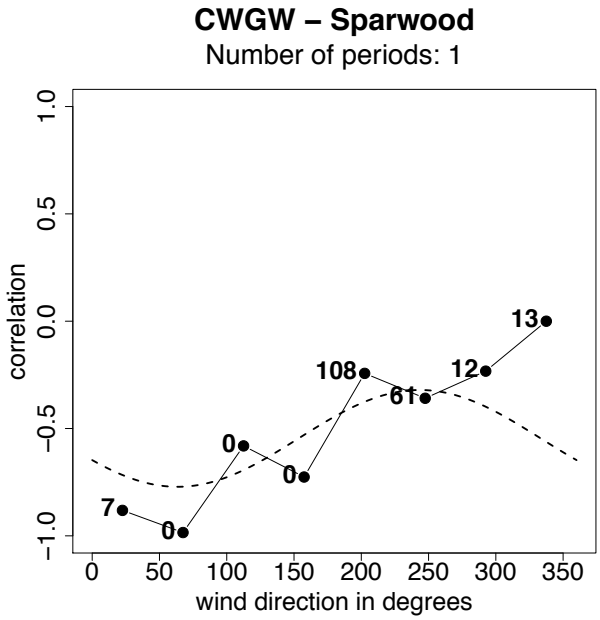
In the current context, we only make forecasts for a single date and a single location and therefore can assume the forecast errors to be independent in space and time. Many applications, however, require wind field forecasts, for which estimates of the spatial and temporal dependencies are needed. Berrocal et al. (2007) combine ensemble forecasts for weather fields with the BMA postprocessing method, which can also be done with both the univariate and bivariate EMOS techniques.

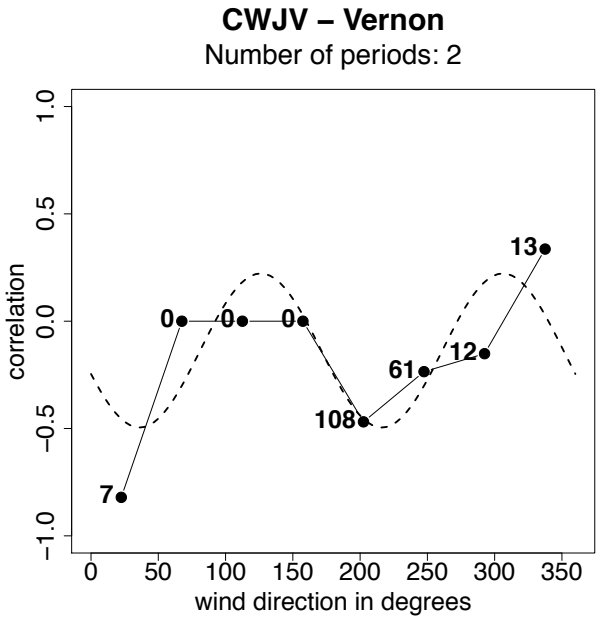
To conclude this thesis, we summarise that the newly developed bivariate EMOS method provides an excellent way to perform statistical postprocessing for two-dimensional weather quantities. It yielded very good results for the UWME, due to modelling the spatial correlation as a function of the predicted wind direction. As the essential properties of the univariate approach are retained, the two-dimensional EMOS technique can be seen as a direct extension.

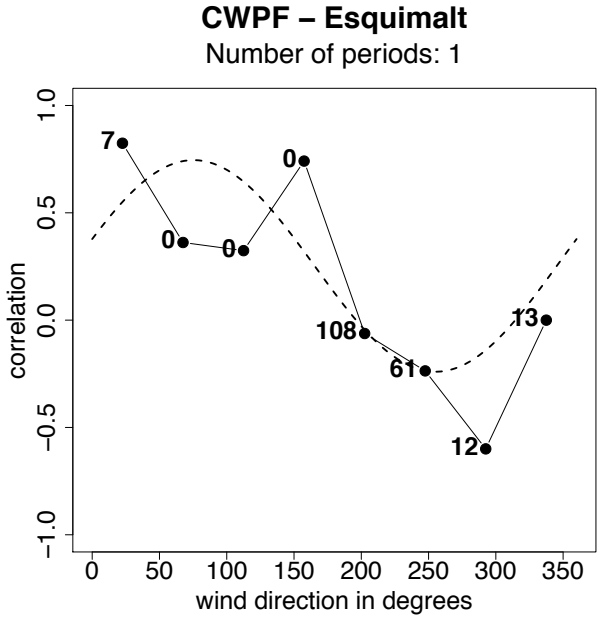
Appendix A

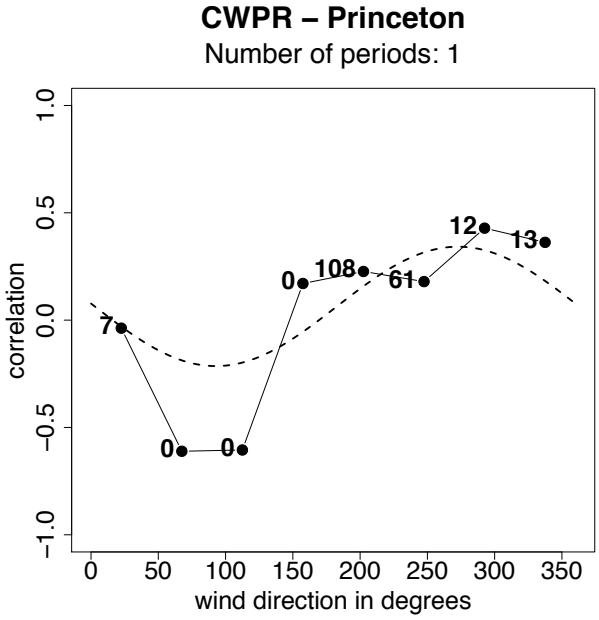
Local Correlation Curves

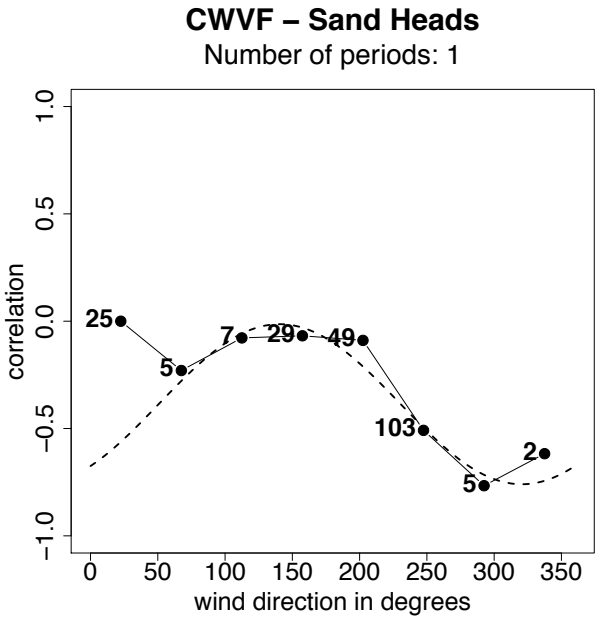
An important feature of the new bivariate EMOS method is the estimation of the correlation structure based on past observations. In Section 4.2.4, we modeled the predictive correlation as a function of the bias-corrected predictive mean vector. Figure 4.8 shows the empirical and estimated correlation curves to be used in the regional EMOS technique, with data aggregated over all observational locations in the North American Pacific Northwest. For the local technique, we create location-specific curves in order to obtain more accurate localised forecasts. These curves were produced using data from the calendar year 2007 and can be found in the following.

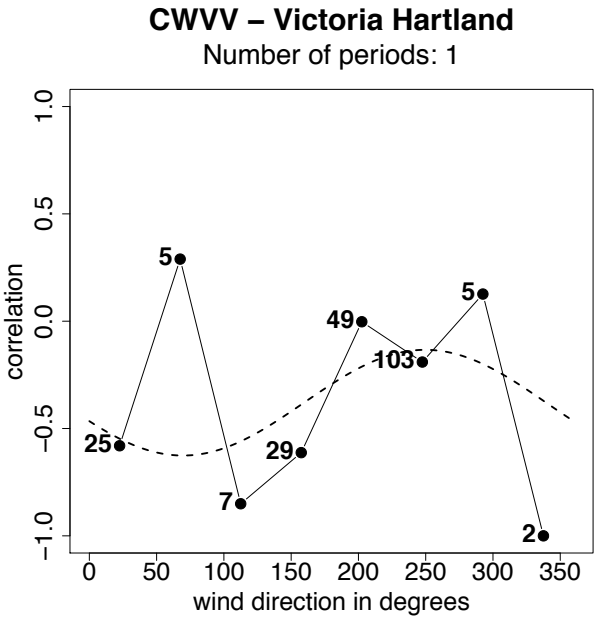

CWGW – Sparwood
Number of periods: 1
correlation
wind direction in degrees
7
0
0
0
108
61
12
13

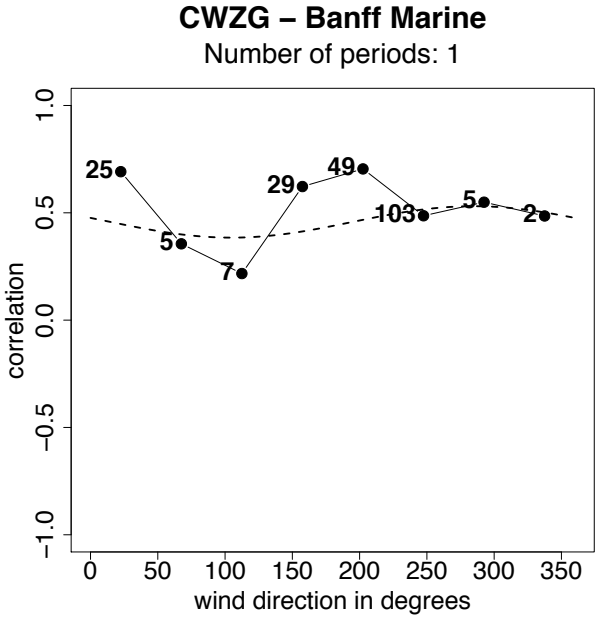

CWJV – Vernon
Number of periods: 2
correlation
wind direction in degrees
7
0
0
0
108
61
12
13

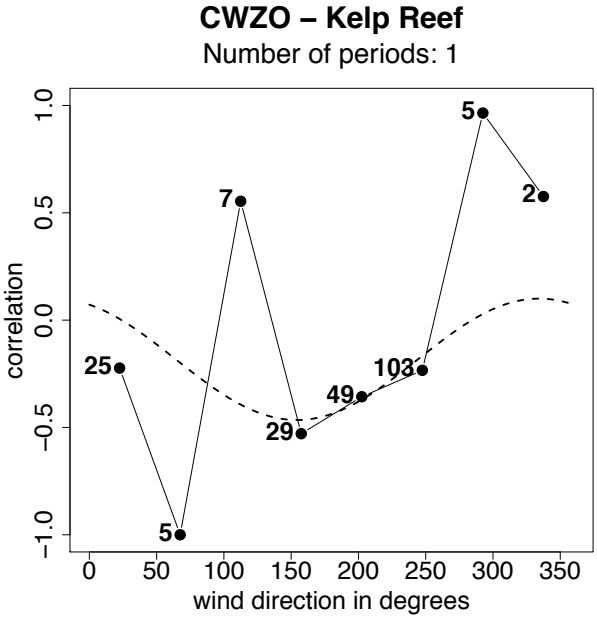

CWPF – Esquimalt
Number of periods: 1
correlation
wind direction in degrees
7
0
0
0
108
61
12
13

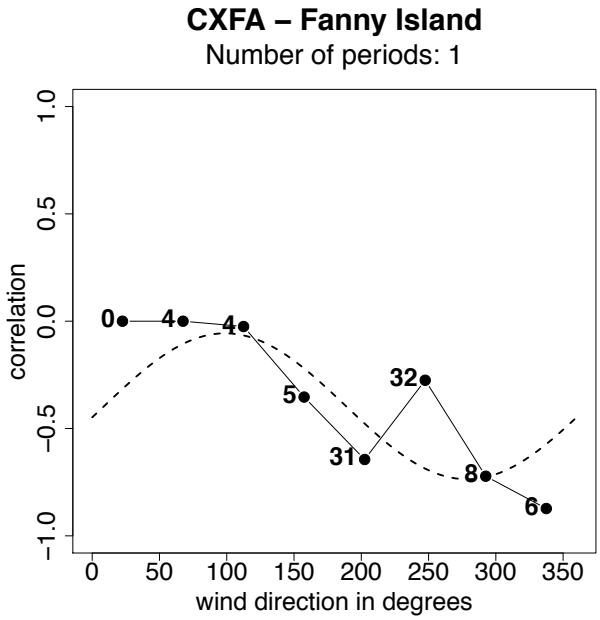

CWPR – Princeton
Number of periods: 1
correlation
wind direction in degrees
7
0
0
0
108
61
12
13

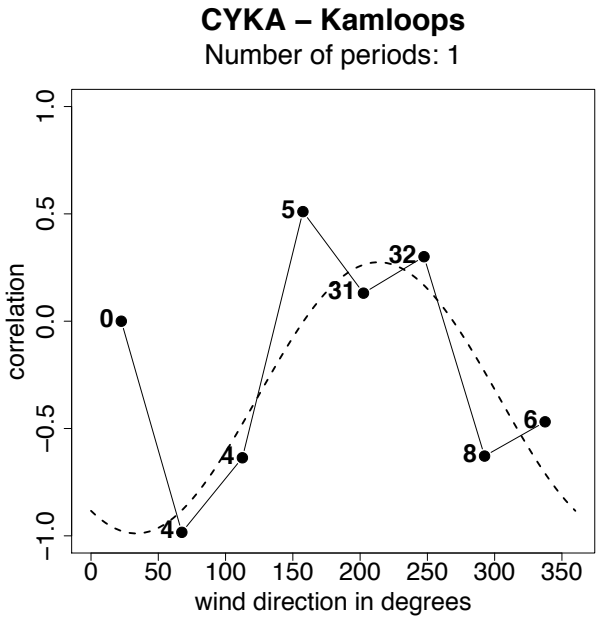

CWVF – Sand Heads
Number of periods: 1
correlation
wind direction in degrees
25
5
7
29
49
103
5
2

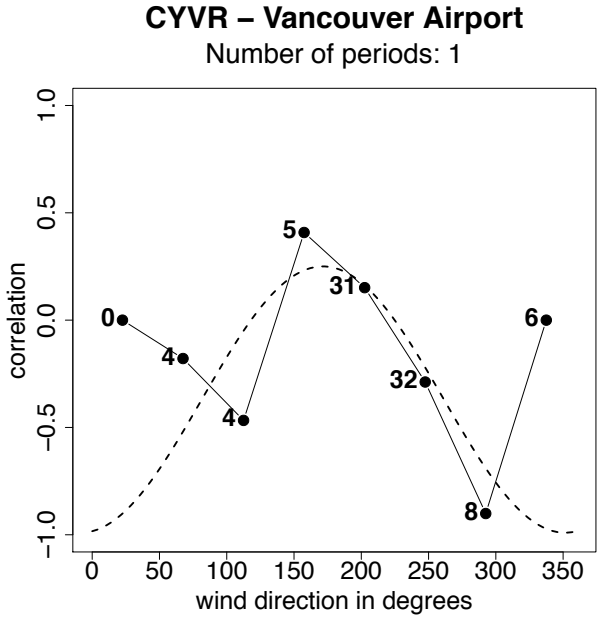

CWVV – Victoria Hartland
Number of periods: 1
correlation
wind direction in degrees
5
49
5
103
25
29
7
2

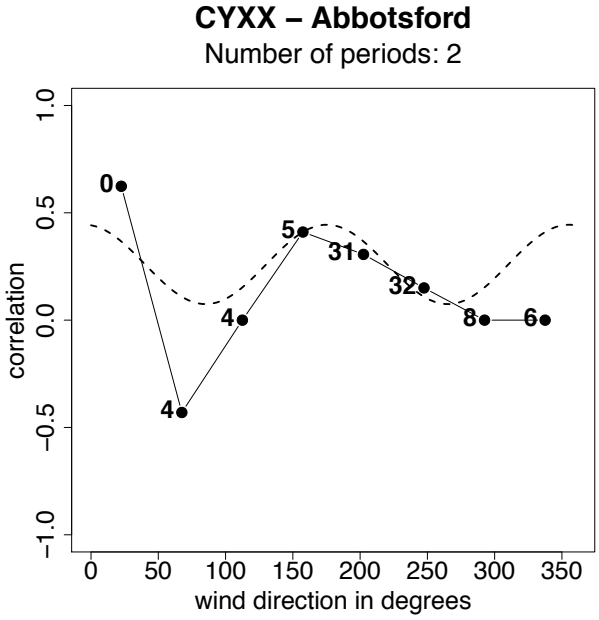

CWZG – Banff Marine
Number of periods: 1
correlation
wind direction in degrees
25
29
49
5
103
5
2
7

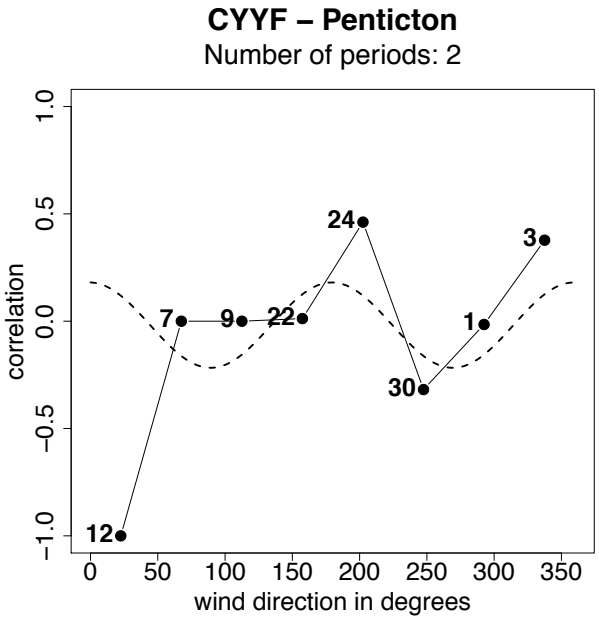

CWZO – Kelp Reef
Number of periods: 1
correlation
wind direction in degrees
5
7
2
25
103
49
29
5

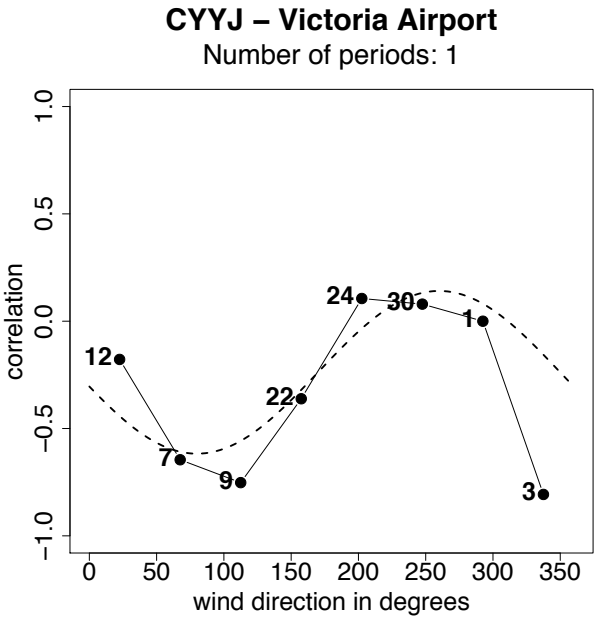

CXFA – Fanny Island
Number of periods: 1
correlation
wind direction in degrees
0
4
4
5
31
32
8
6

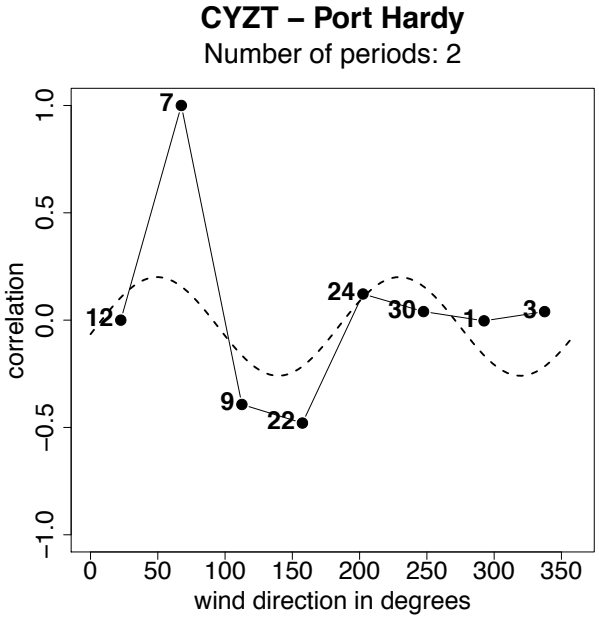

CYKA – Kamloops
Number of periods: 1
correlation
wind direction in degrees
0
4
5
31
32
8
6

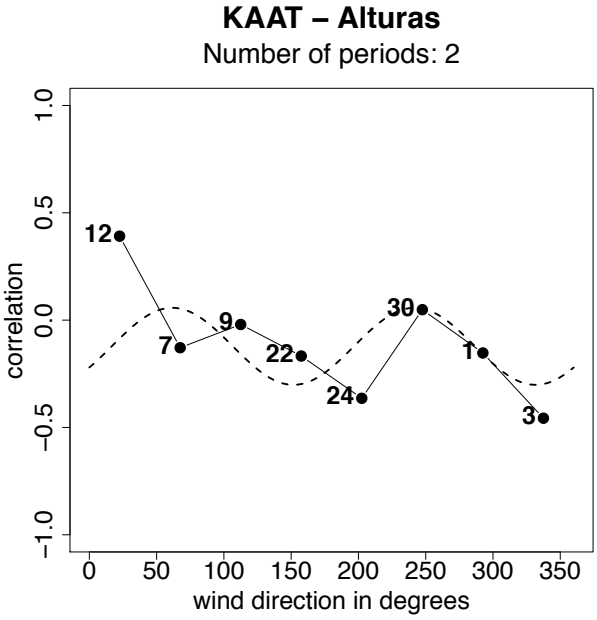

CYVR – Vancouver Airport
Number of periods: 1
correlation
wind direction in degrees
0
4
4
5
31
32
8
6

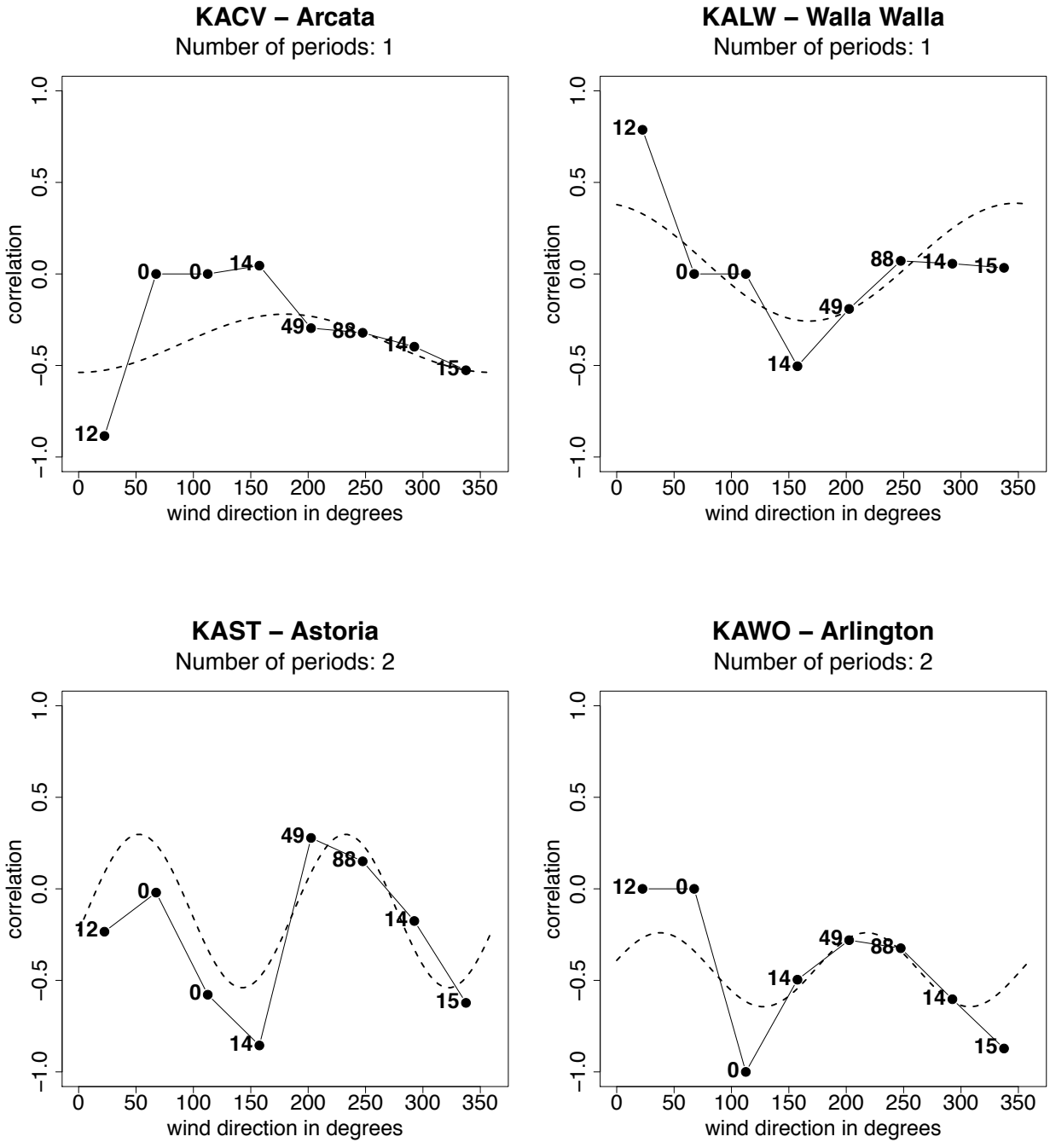

CYXX – Abbotsford
Number of periods: 2
correlation
wind direction in degrees
0
4
5
31
32
8
6

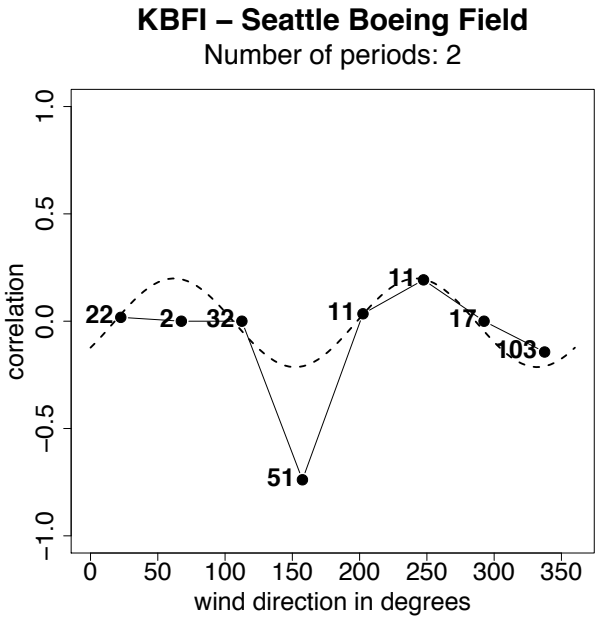

CYYF – Penticton
Number of periods: 2
correlation
wind direction in degrees
12
7
9
22
24
30
1
3

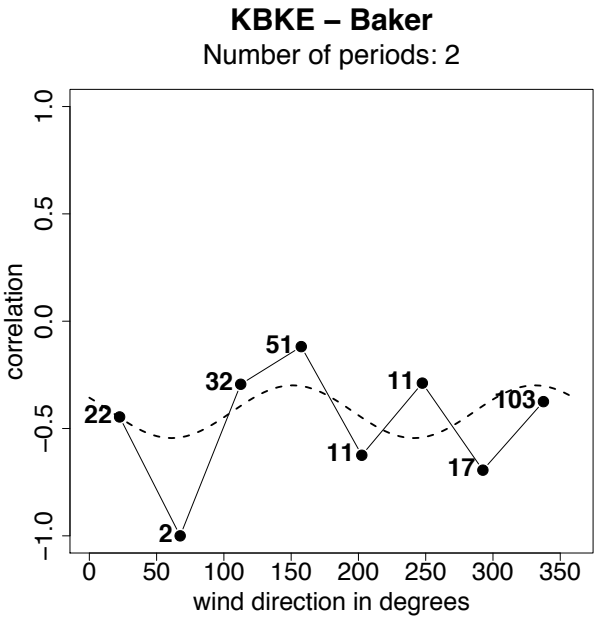

CYYJ – Victoria Airport
Number of periods: 1
correlation
wind direction in degrees
12
7
9
22
24
30
1
3

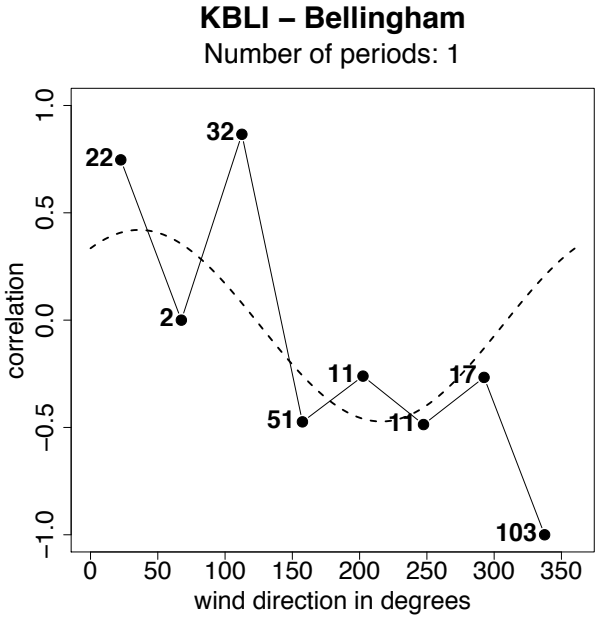

CYZT – Port Hardy
Number of periods: 2
correlation
wind direction in degrees
12
7
9
22
24
30
1
3

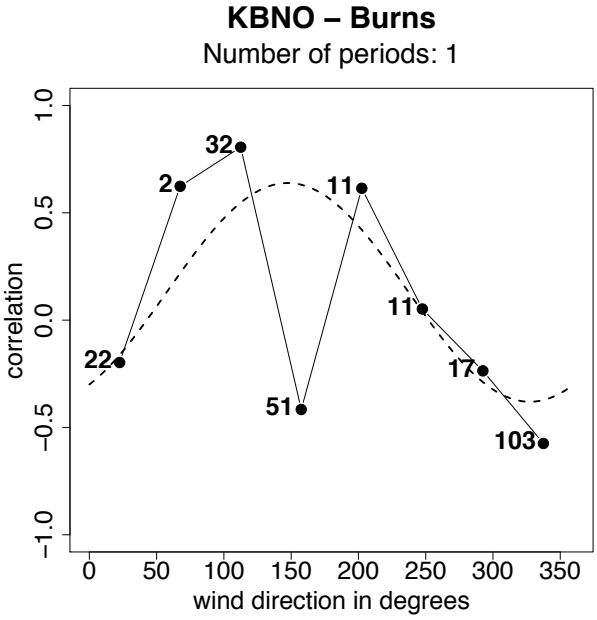

KAAT – Alturas
Number of periods: 2
correlation
wind direction in degrees
12
7
9
22
24
30
1
3

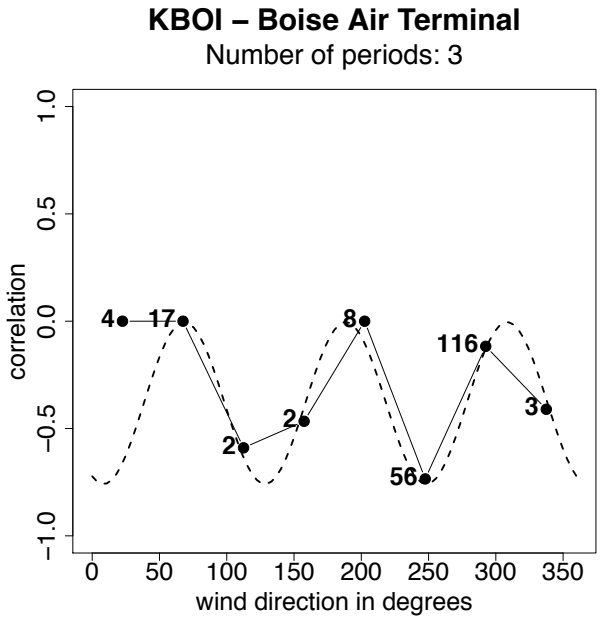

KACV – Arcata
Number of periods: 1
KALW – Walla Walla
Number of periods: 1
KAST – Astoria
Number of periods: 2
KAWO – Arlington
Number of periods: 2
correlation
wind direction in degrees
−1.0
−0.5
0.0
0.5
1.0
0
50
100
150
200
250
300
350
12
0
14
49
88
15

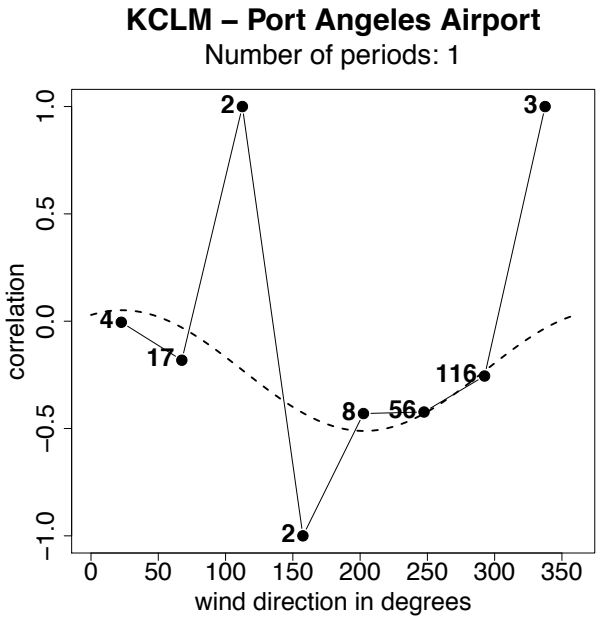

KBFI – Seattle Boeing Field
Number of periods: 2
correlation
wind direction in degrees
22
2
32
51
11
11
17
103

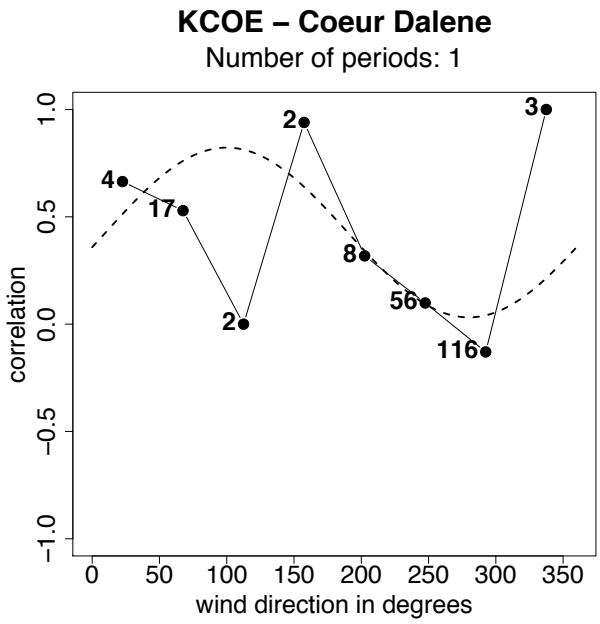

KBKE – Baker
Number of periods: 2
correlation
wind direction in degrees
22
2
32
51
11
11
17
103

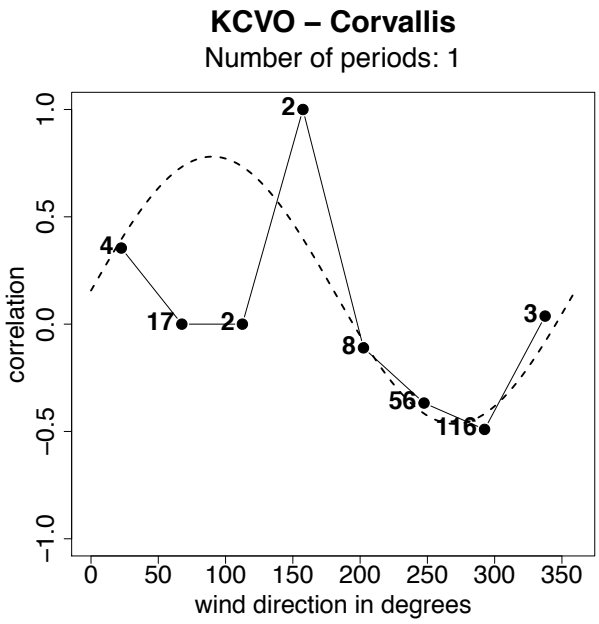

KBLI – Bellingham
Number of periods: 1
correlation
wind direction in degrees
22
2
32
51
11
11
17
103

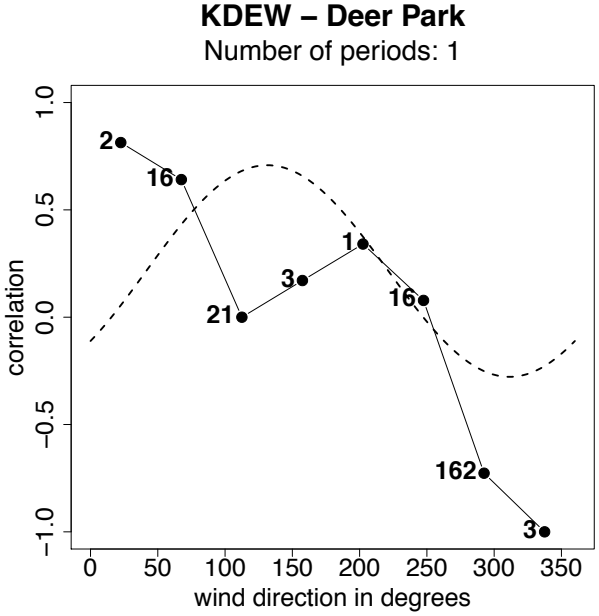

KBNO – Burns
Number of periods: 1
correlation
wind direction in degrees
22
2
32
51
11
11
17
103

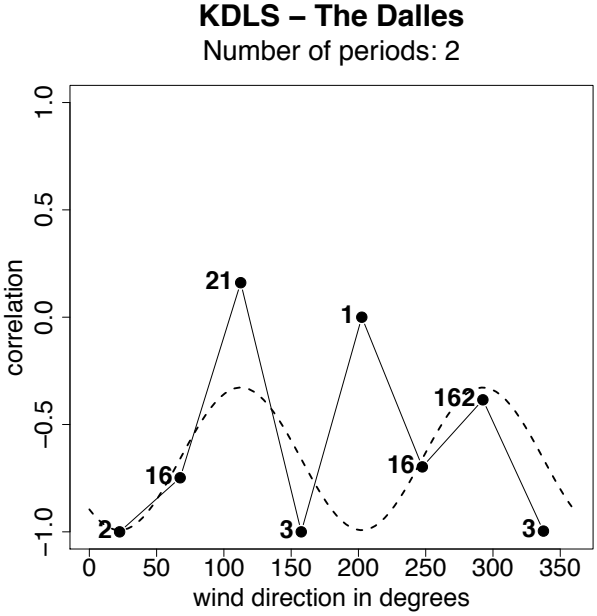

KBOI – Boise Air Terminal
Number of periods: 3
correlation
wind direction in degrees
4
17
2
2
8
56
116
3

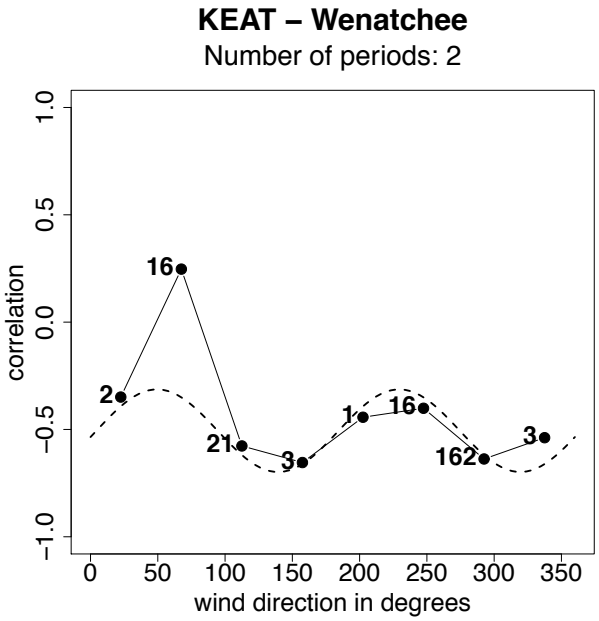

KCLM – Port Angeles Airport
Number of periods: 1
correlation
wind direction in degrees
4
17
2
2
8
56
116
3

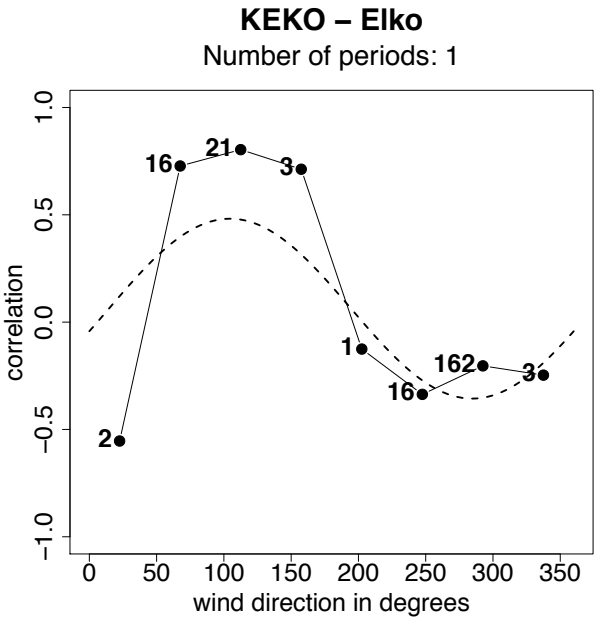

KCOE – Coeur Dalene
Number of periods: 1
correlation
wind direction in degrees
4
17
2
2
8
56
116
3

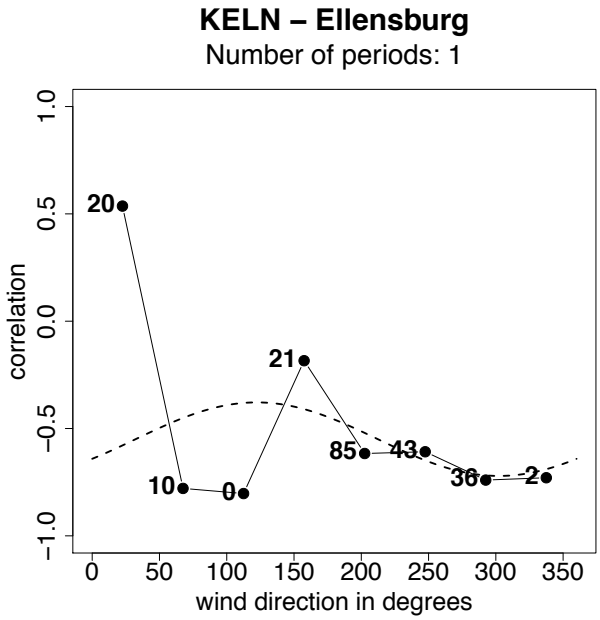

KCVO – Corvallis
Number of periods: 1
correlation
wind direction in degrees
4
17
2
2
8
56
116
3

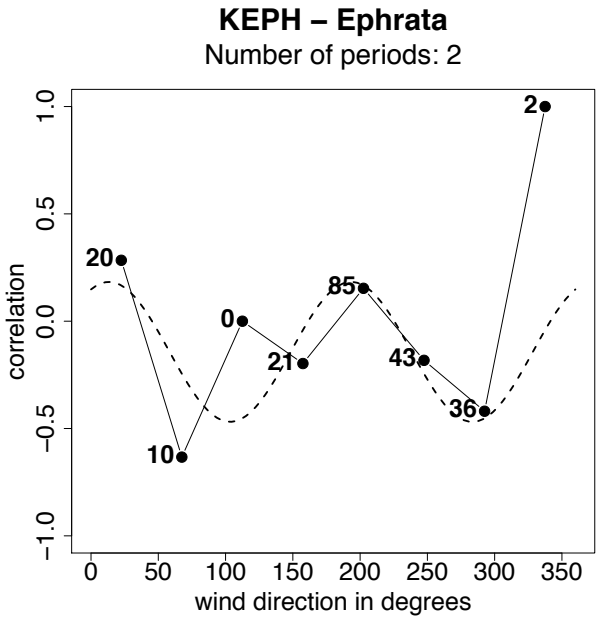

KDEW – Deer Park
Number of periods: 1
correlation
wind direction in degrees
2
16
21
3
1
16
162
3

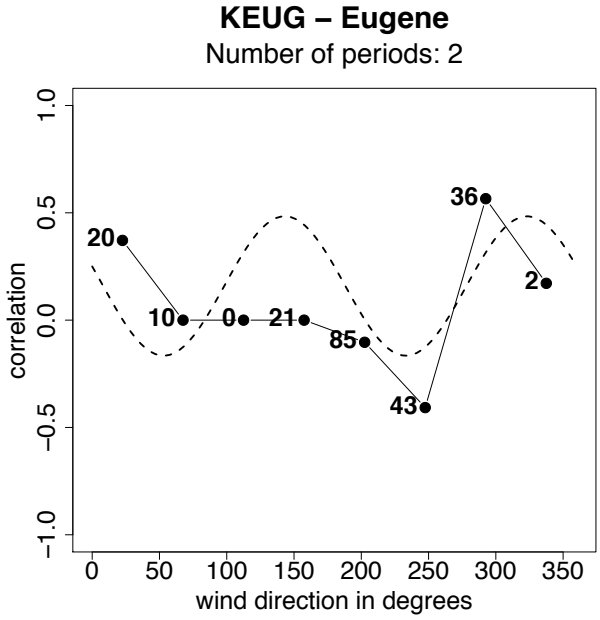

KDLS – The Dalles
Number of periods: 2
correlation
wind direction in degrees
21
1
162
16
16
2
3
3

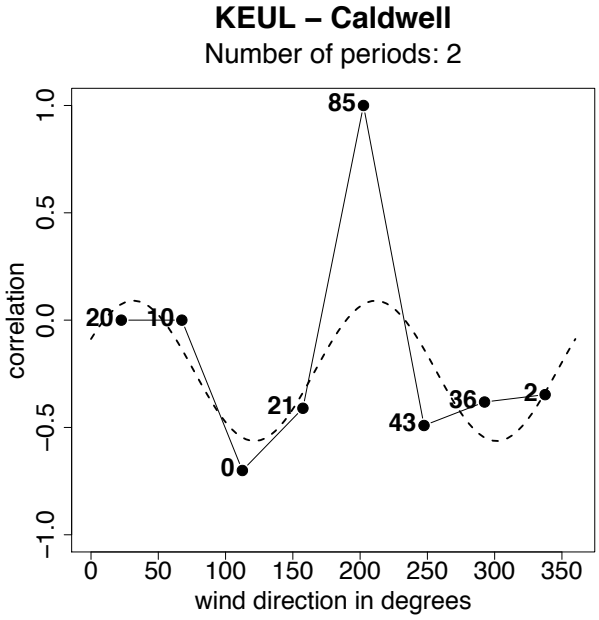

KEAT – Wenatchee
Number of periods: 2
correlation
wind direction in degrees
16
2
21
3
1
16
162
3

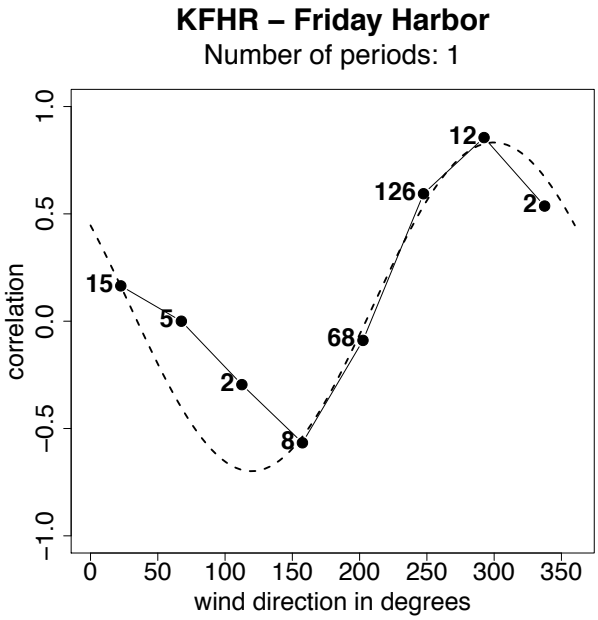

KEKO – Elko
Number of periods: 1
correlation
wind direction in degrees
16
21
3
1
162
16
3
2

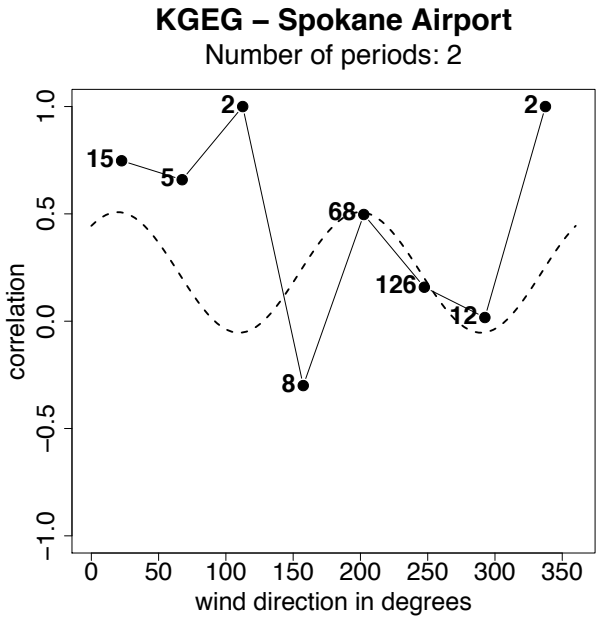

KELN – Ellensburg
Number of periods: 1
correlation
wind direction in degrees
20
10
0
21
85
43
36
2

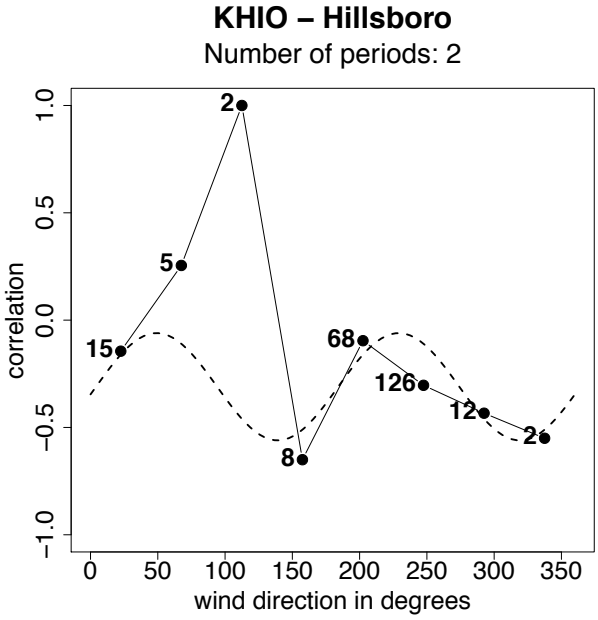

KEPH – Ephrata
Number of periods: 2
correlation
wind direction in degrees
20
10
0
21
85
43
36
2

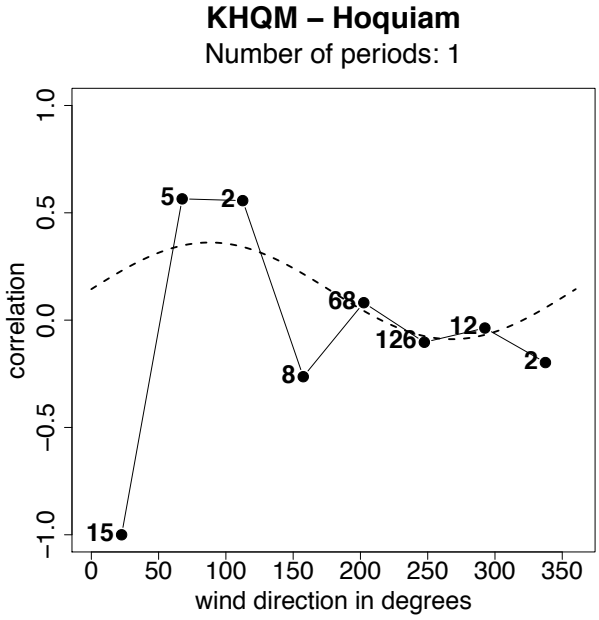

KEUG – Eugene
Number of periods: 2
correlation
wind direction in degrees
20
10
0
21
85
43
36
2

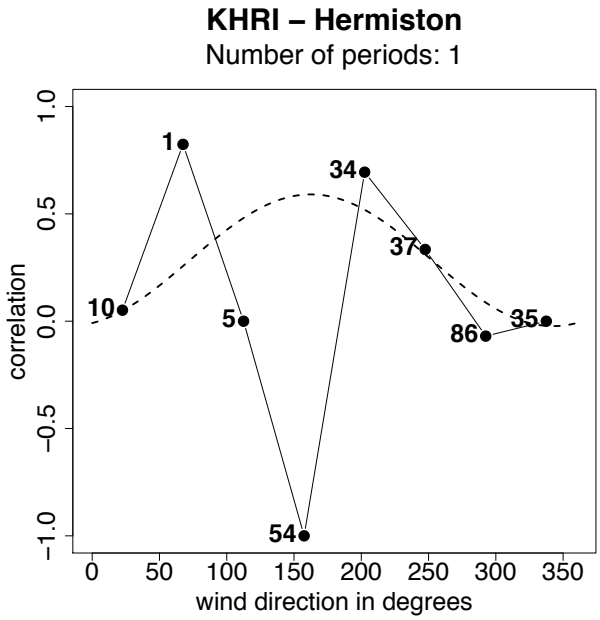

KEUL – Caldwell
Number of periods: 2
correlation
wind direction in degrees
20
10
0
21
85
43
36
2

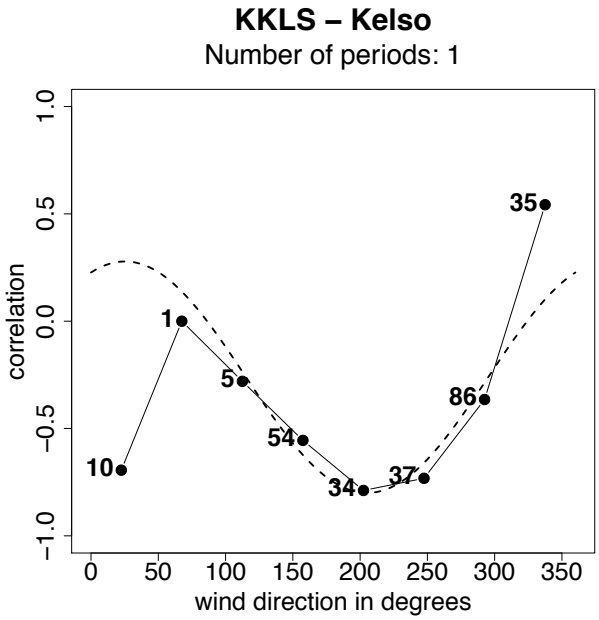

KFHR – Friday Harbor
Number of periods: 1
correlation
wind direction in degrees
15
5
2
8
68
126
12
2

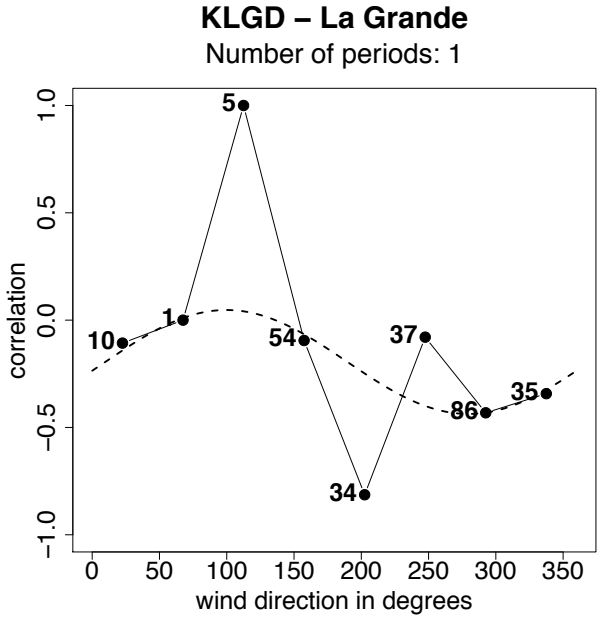

KGEG – Spokane Airport
Number of periods: 2
correlation
wind direction in degrees
15
5
2
8
68
126
12
2

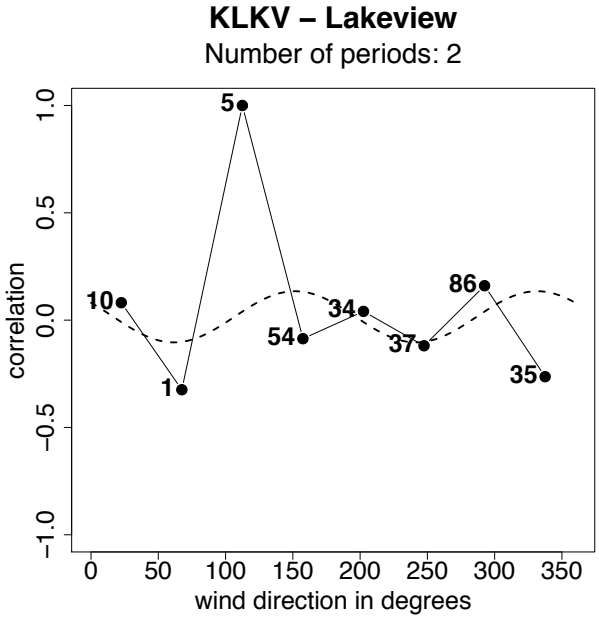

KHIO – Hillsboro
Number of periods: 2
correlation
wind direction in degrees
15
5
2
8
68
126
12
2

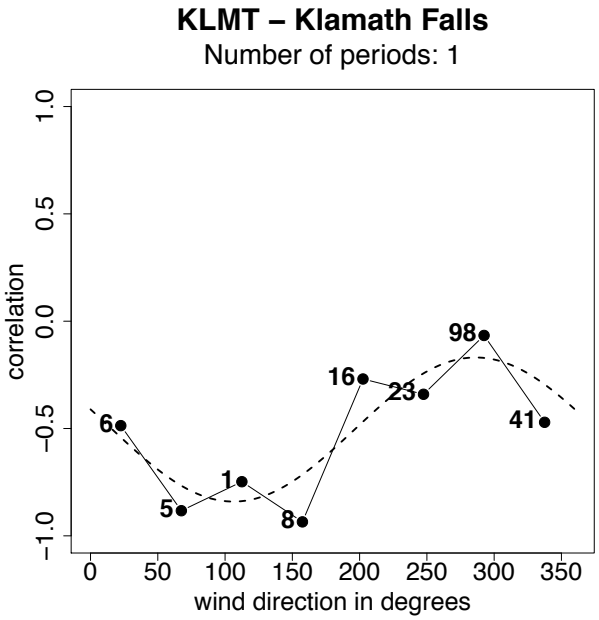

KHQM – Hoquiam
Number of periods: 1
correlation
wind direction in degrees
15
5
2
8
68
126
12
2

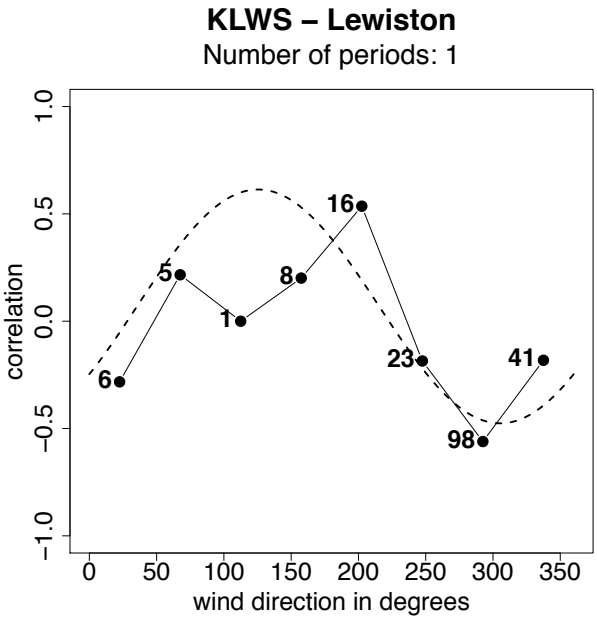

KHRI – Hermiston
Number of periods: 1
correlation
wind direction in degrees
10
1
5
54
34
37
86
35

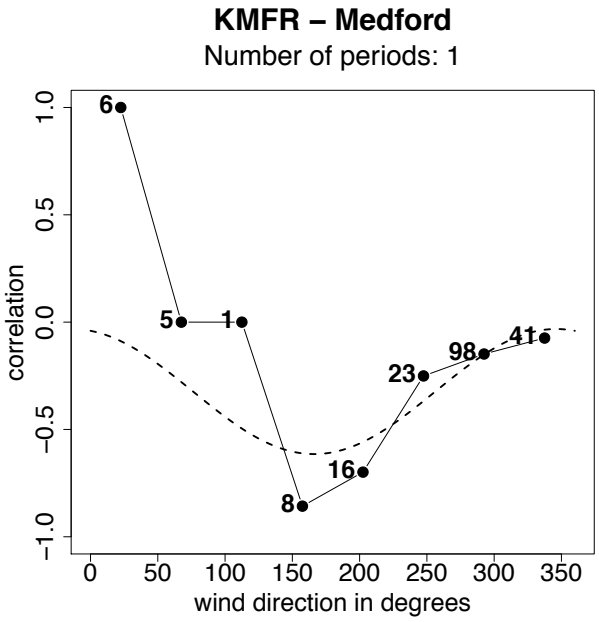

KKLS – Kelso
Number of periods: 1
correlation
wind direction in degrees
10
1
5
54
34
37
86
35

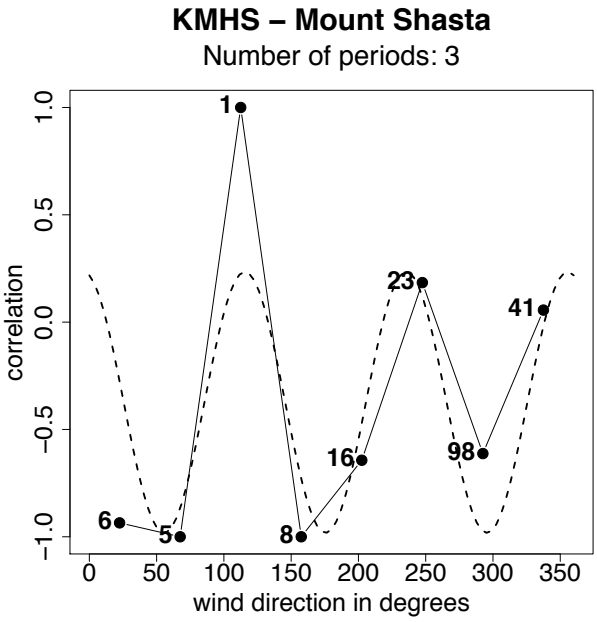

KLGD – La Grande
Number of periods: 1
correlation
wind direction in degrees
10
1
5
54
34
37
86
35

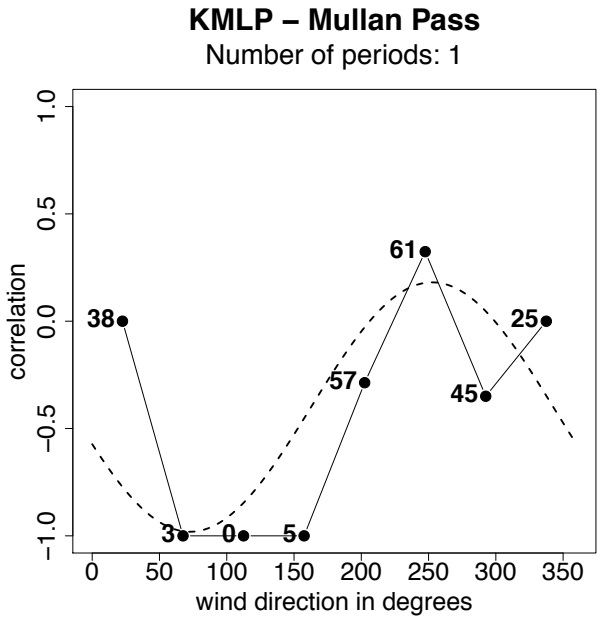

KLKV – Lakeview
Number of periods: 2
correlation
wind direction in degrees
10
1
5
54
34
37
86
35

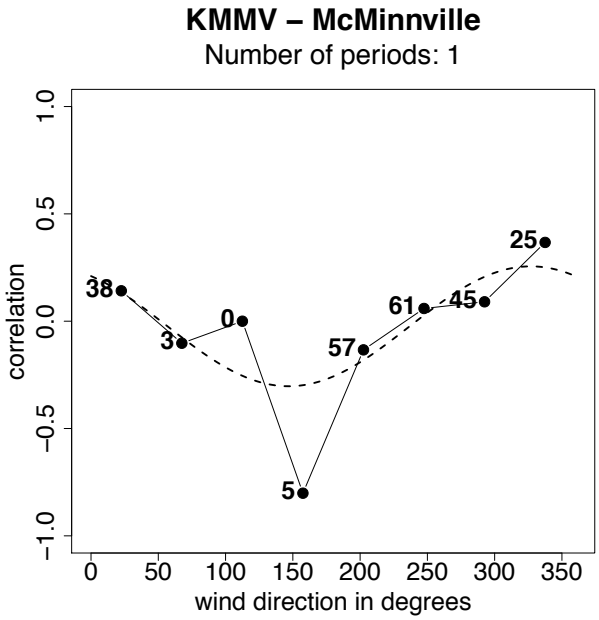

KLMT – Klamath Falls
Number of periods: 1
correlation
wind direction in degrees
6
5
1
8
16
23
98
41

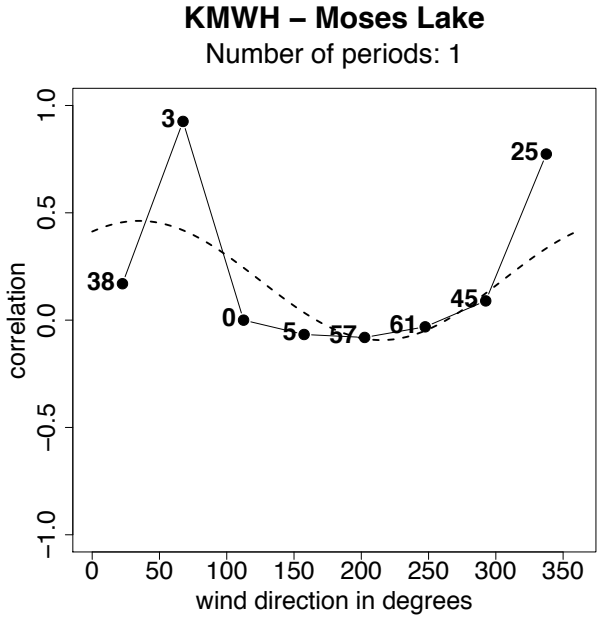

KLWS – Lewiston
Number of periods: 1
correlation
wind direction in degrees
6
5
1
8
16
23
98
41

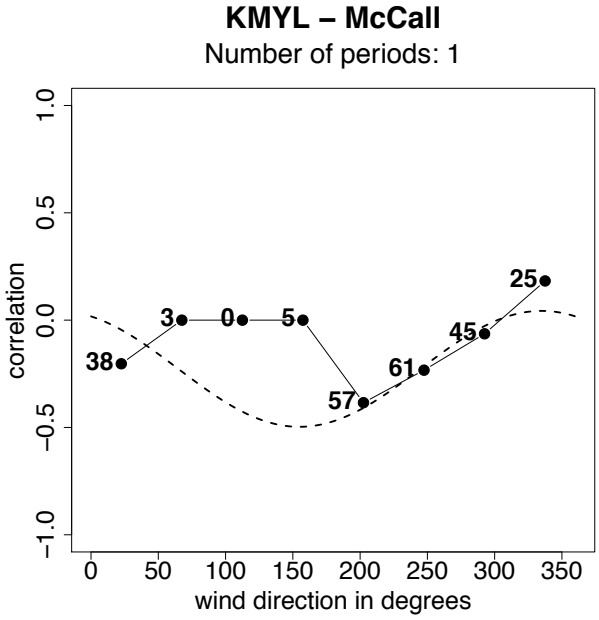

KMFR – Medford
Number of periods: 1
correlation
wind direction in degrees
6
5
1
8
16
23
98
41

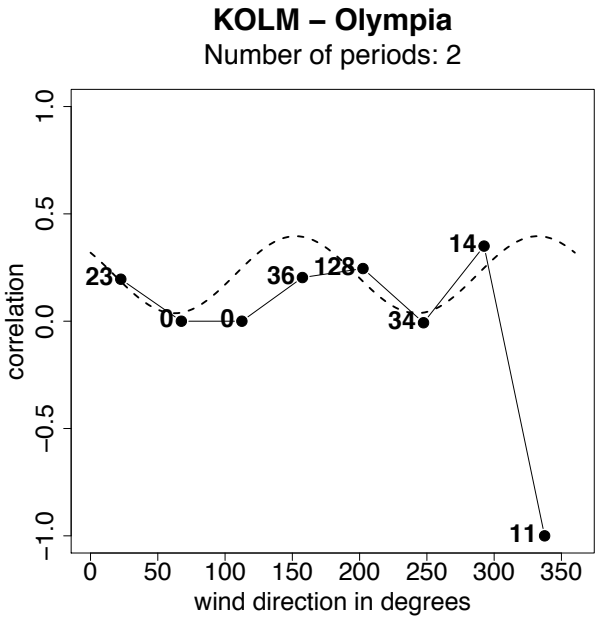

KMHS – Mount Shasta
Number of periods: 3
correlation
wind direction in degrees
1
6
5
8
16
23
98
41

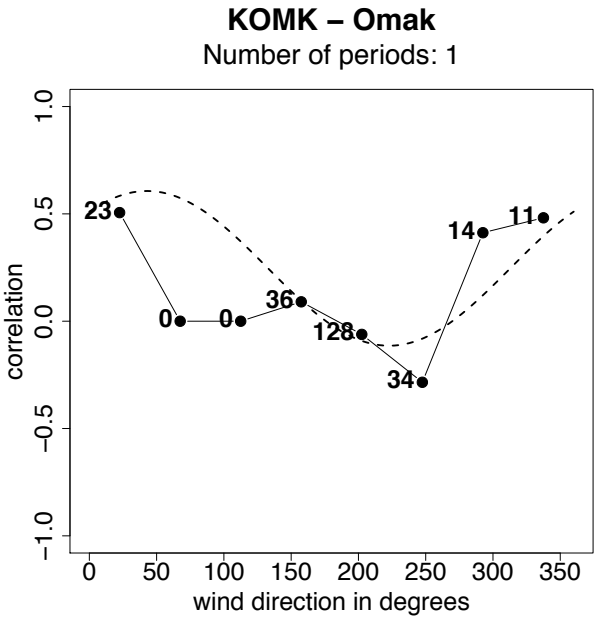

KMLP – Mullan Pass
Number of periods: 1
correlation
wind direction in degrees
38
3
0
5
57
61
45
25

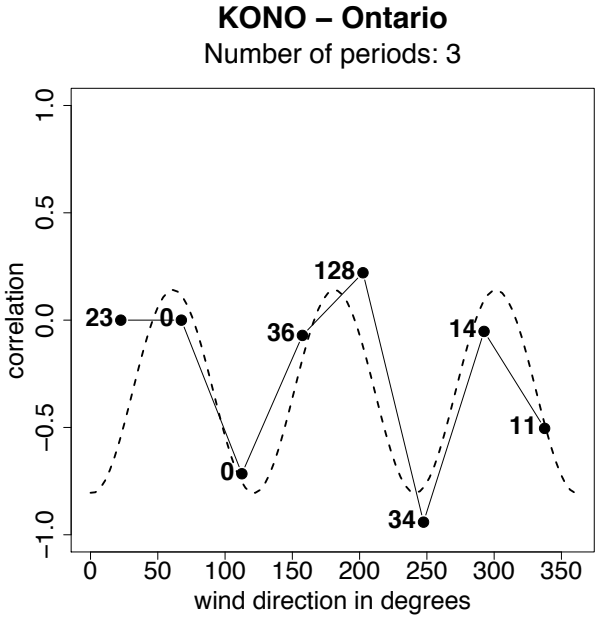

KMMV – McMinnville
Number of periods: 1
correlation
wind direction in degrees
38
3
0
5
57
61
45
25

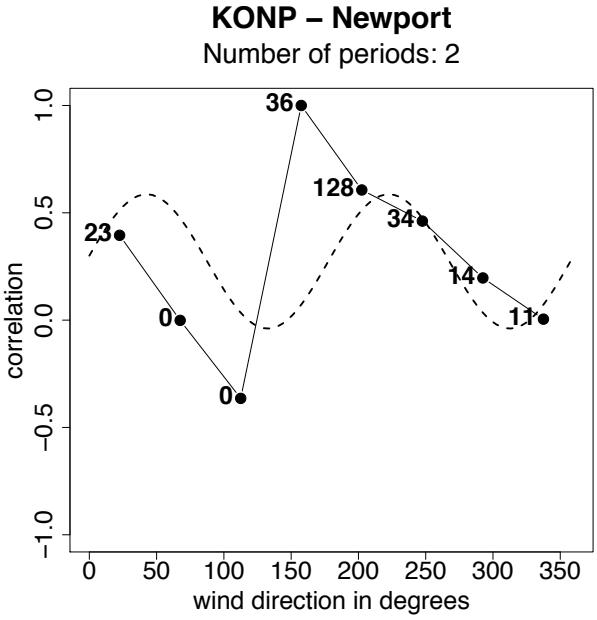

KMWH – Moses Lake
Number of periods: 1
correlation
wind direction in degrees
38
3
0
5
57
61
45
25

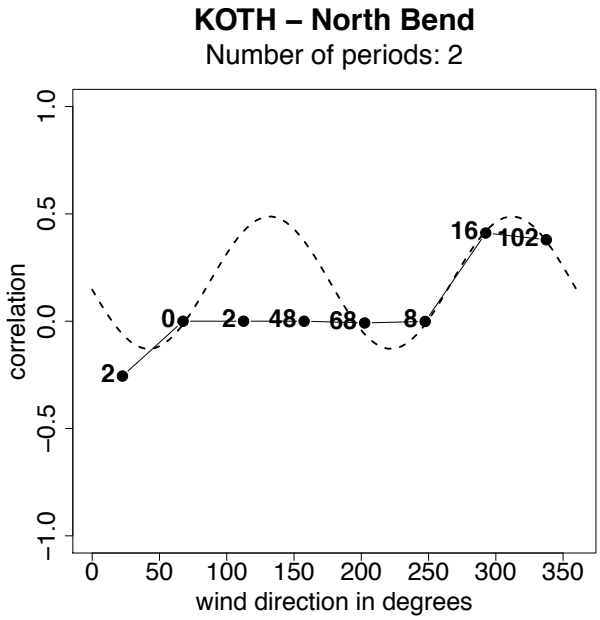

KMYL – McCall
Number of periods: 1
correlation
wind direction in degrees
38
3
0
5
57
61
45
25

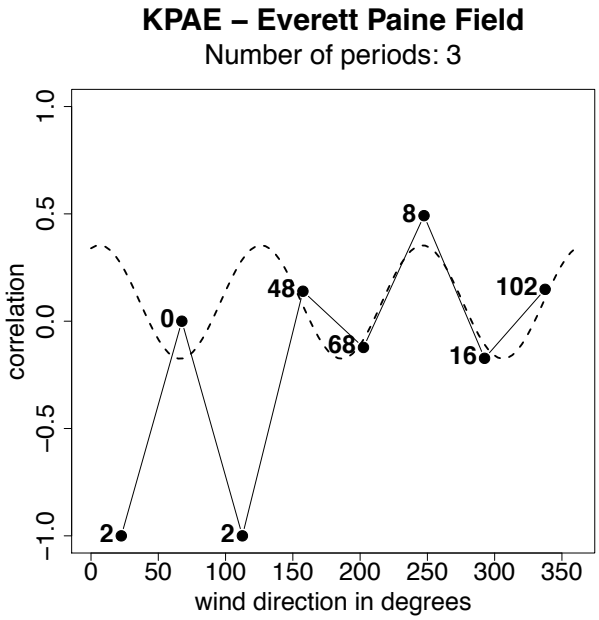

KOLM – Olympia
Number of periods: 2
correlation
wind direction in degrees
23
0
0
36
128
34
14
11

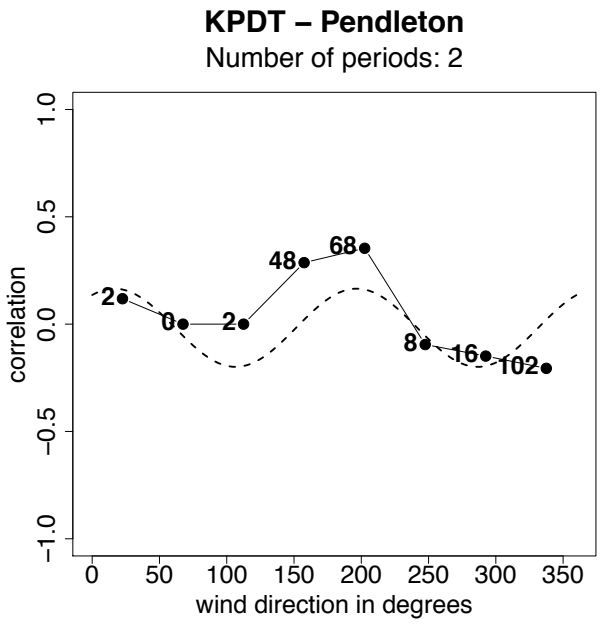

KOMK – Omak
Number of periods: 1
correlation
wind direction in degrees
23
0
0
36
128
34
14
11

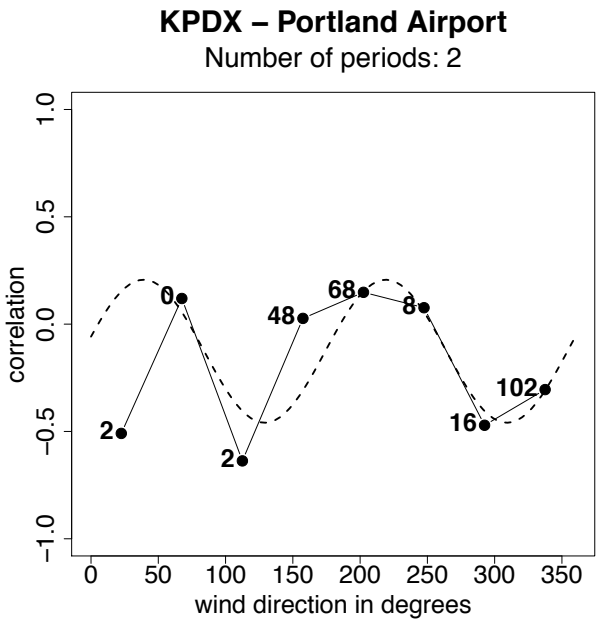

KONO – Ontario
Number of periods: 3
correlation
wind direction in degrees
23
0
36
128
0
34
14
11

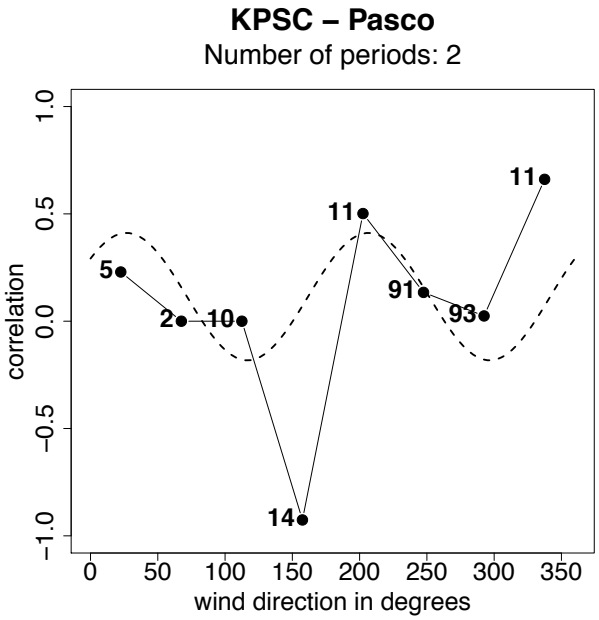

KONP – Newport
Number of periods: 2
correlation
wind direction in degrees
23
0
36
0
128
34
14
11

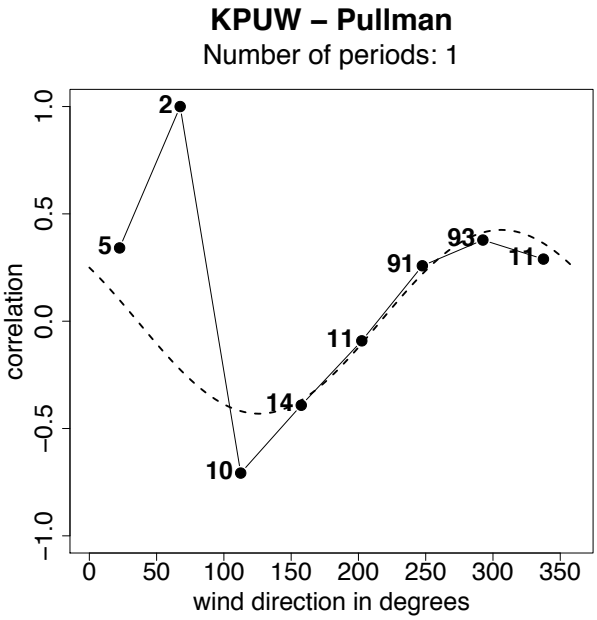

KOTH – North Bend
Number of periods: 2
correlation
wind direction in degrees
2
0
2
48
68
8
16
102

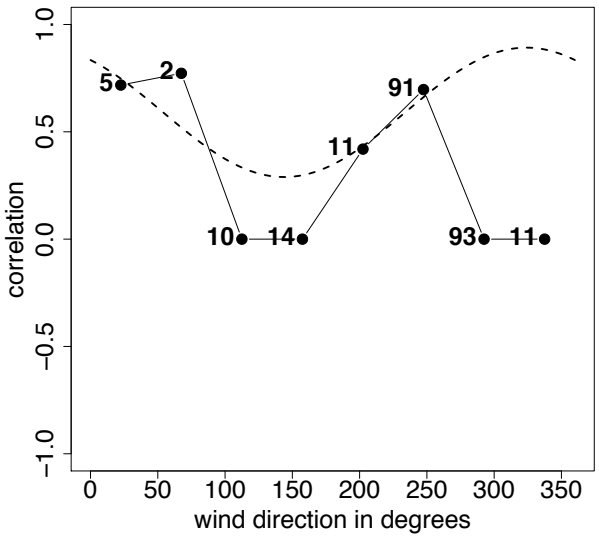

KPAE – Everett Paine Field
Number of periods: 3
correlation
wind direction in degrees
2
0
2
48
68
8
16
102

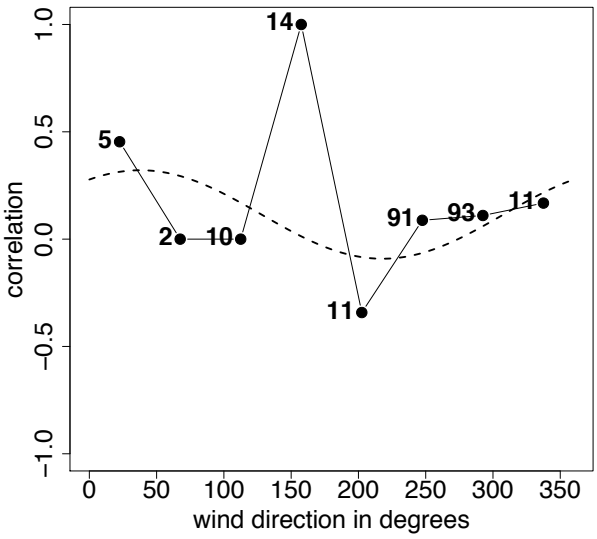

KPDT – Pendleton
Number of periods: 2
correlation
wind direction in degrees
2
0
2
48
68
8
16
102

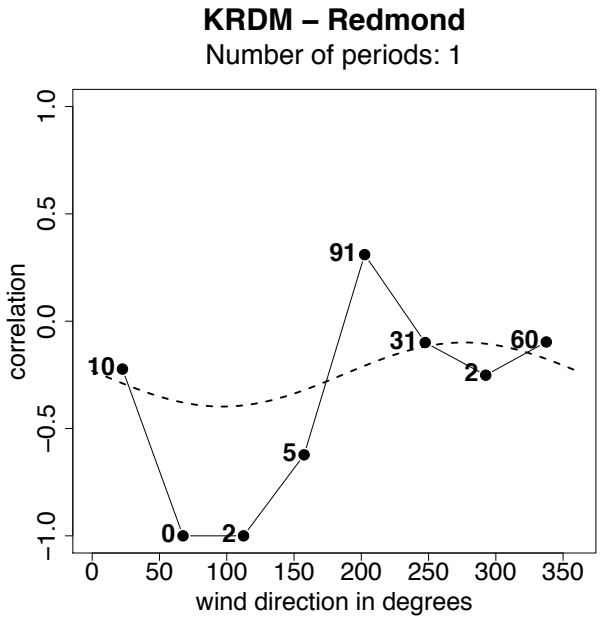

KPDX – Portland Airport
Number of periods: 2
correlation
wind direction in degrees
2
0
2
48
68
8
16
102

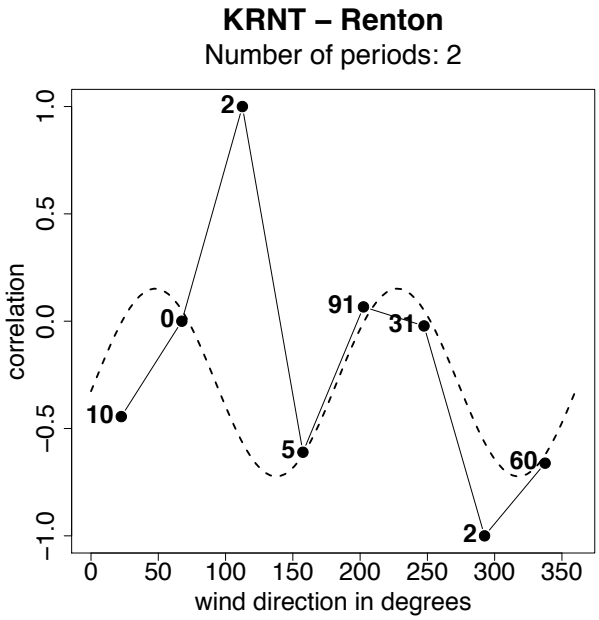

KPSC – Pasco
Number of periods: 2
correlation
wind direction in degrees
5
2
10
14
11
11
91
93
11

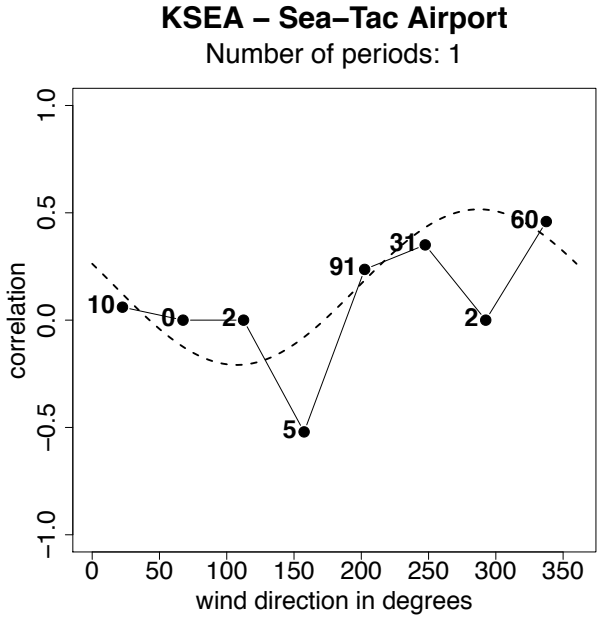

KPUW – Pullman
Number of periods: 1
correlation
wind direction in degrees
2
5
91
93
11
11
14
10

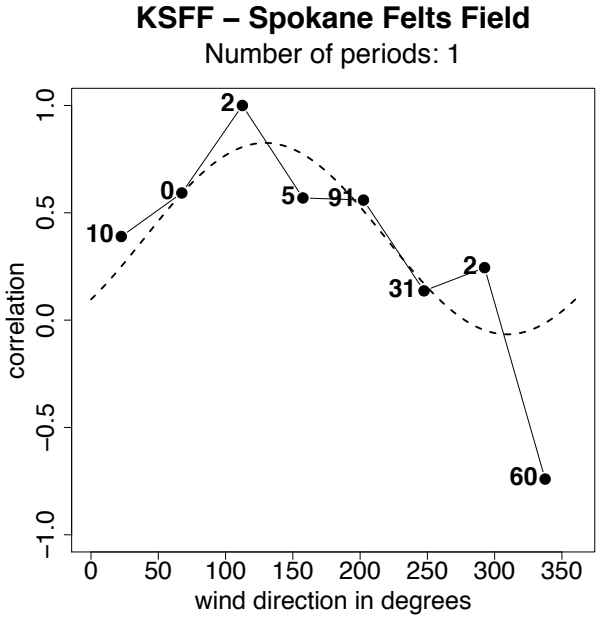

KPWT – Bremerton Airport
Number of periods: 1
correlation
wind direction in degrees
5
2
91
11
10
14
93
11

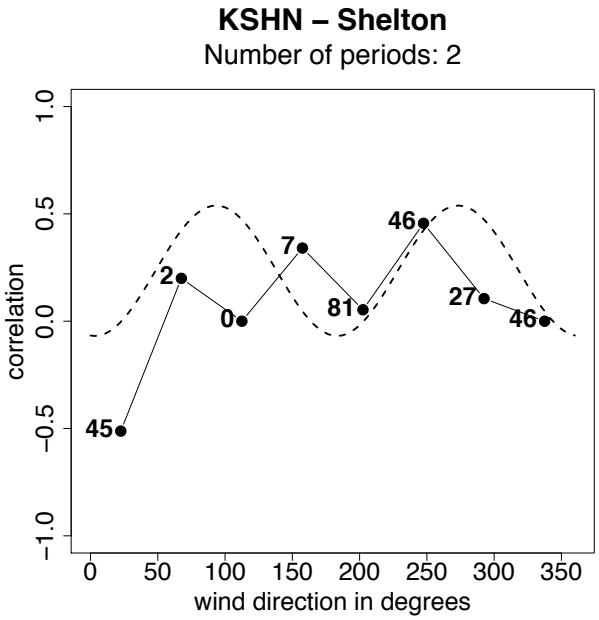

KRBG – Roseburg
Number of periods: 1
correlation
wind direction in degrees
14
5
2
10
11
91
93
11

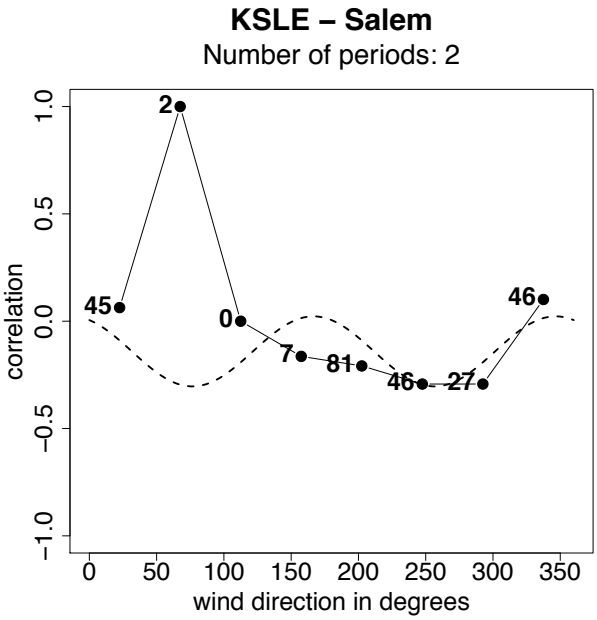

KRDM – Redmond
Number of periods: 1
correlation
wind direction in degrees
10
0
2
5
91
31
2
60

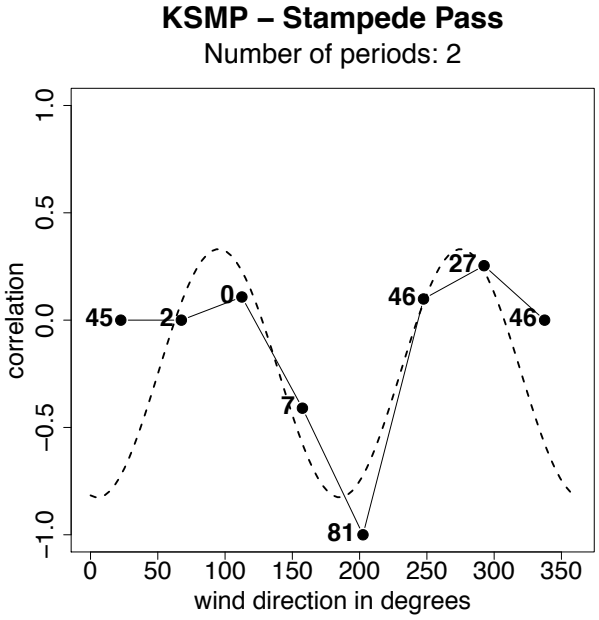

KRNT – Renton
Number of periods: 2
correlation
wind direction in degrees
10
0
2
5
91
31
2
60

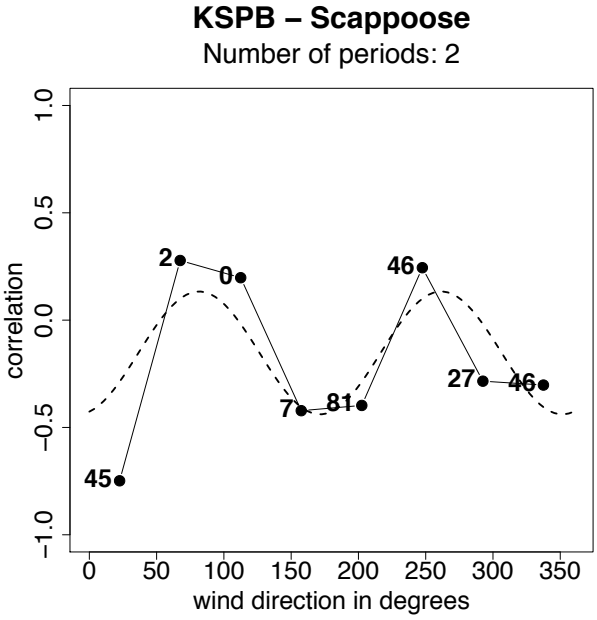

KSEA – Sea–Tac Airport
Number of periods: 1
correlation
wind direction in degrees
10
0
2
5
91
31
2
60

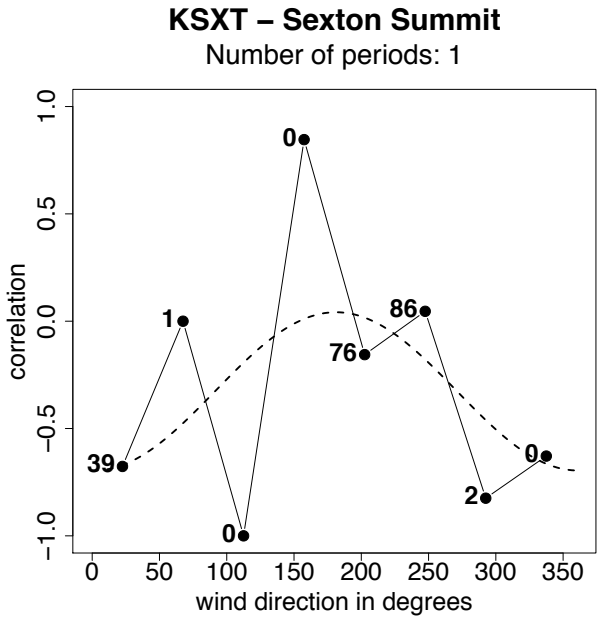

KSFF – Spokane Felts Field
Number of periods: 1
correlation
wind direction in degrees
10
0
2
5
91
31
2
60

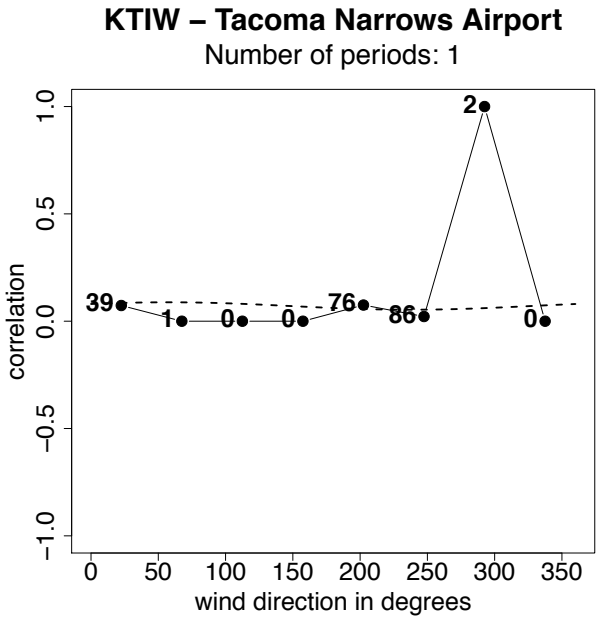

KSHN – Shelton
Number of periods: 2
correlation
wind direction in degrees
45
2
0
7
81
46
27
46

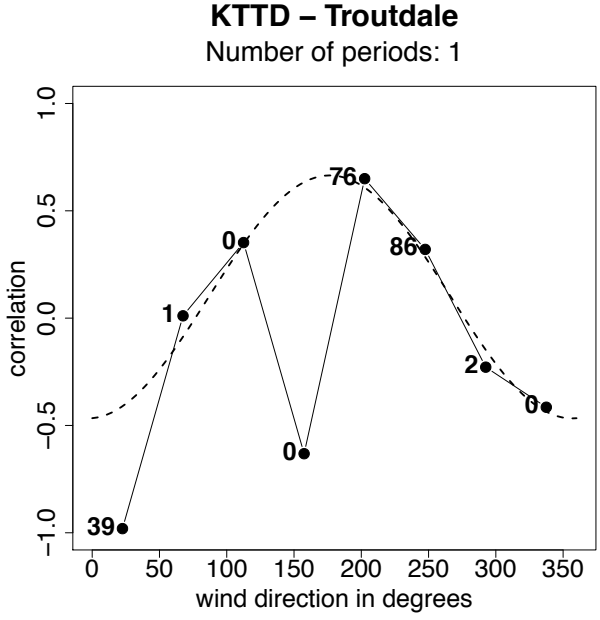

KSLE – Salem
Number of periods: 2
correlation
wind direction in degrees
45
2
0
7
81
46
27
46

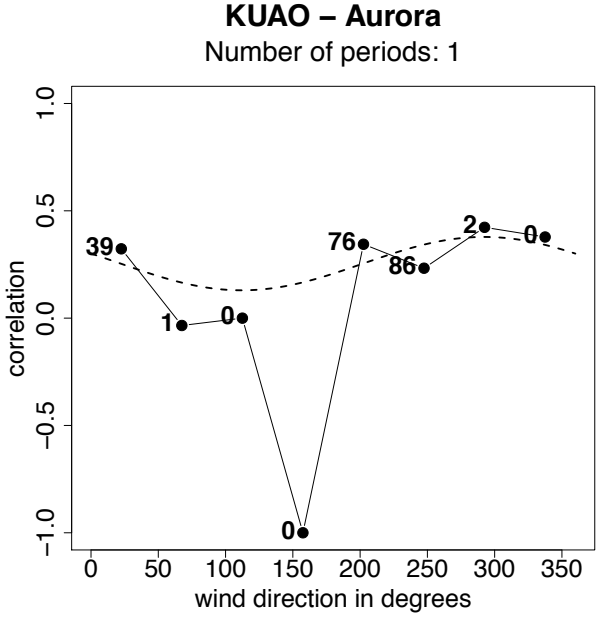

KSMP – Stampede Pass
Number of periods: 2
correlation
wind direction in degrees
45
2
0
7
81
46
27
46

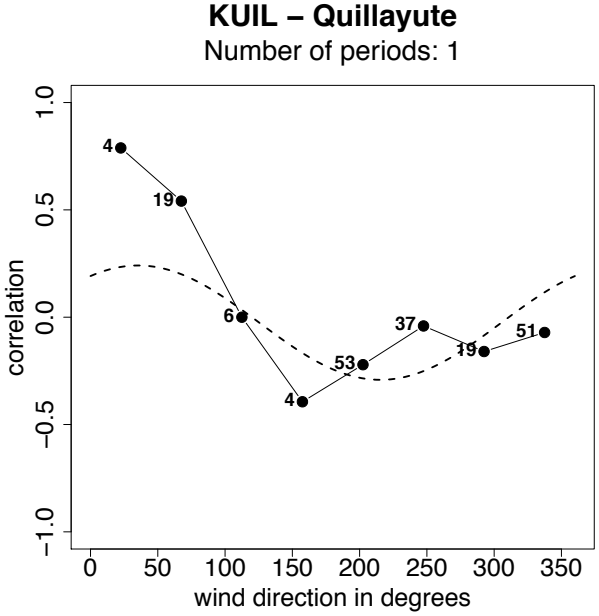

KSPB – Scappoose
Number of periods: 2
correlation
wind direction in degrees
45
2
0
7
81
46
27
46

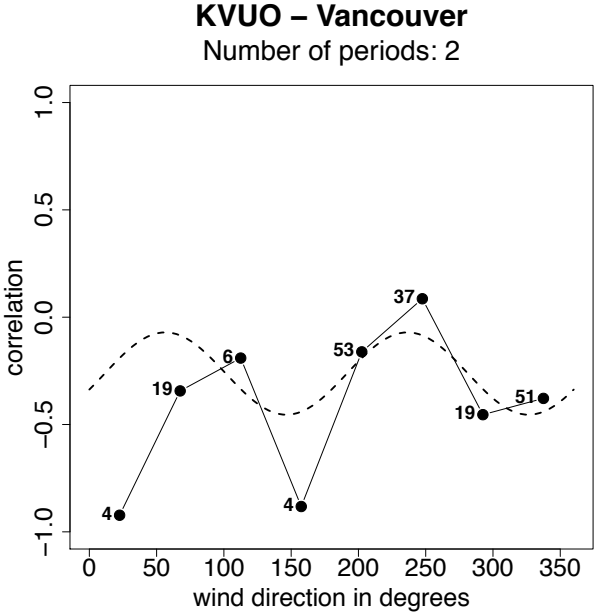

KSXT – Sexton Summit
Number of periods: 1
correlation
wind direction in degrees
39
1
0
76
86
0
2
0

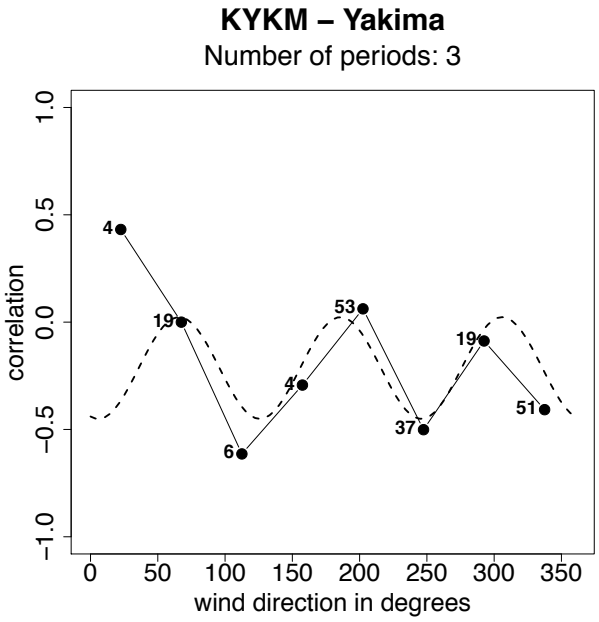

KTIW – Tacoma Narrows Airport
Number of periods: 1
correlation
wind direction in degrees
39
1
0
0
76
86
2
0


KTTD – Troutdale
Number of periods: 1
correlation
wind direction in degrees
39
1
0
76
86
0
2
0


KUAO – Aurora
Number of periods: 1
correlation
wind direction in degrees
39
1
0
0
76
86
2
0


KUIL – Quillayute
Number of periods: 1
correlation
wind direction in degrees
4
19
6
4
53
37
19
51


KVUO – Vancouver
Number of periods: 2
correlation
wind direction in degrees
4
19
6
4
53
37
19
51


KYKM – Yakima
Number of periods: 3
correlation
wind direction in degrees
4
19
6
4
53
37
19
51

Appendix B

List of Notation for Chapter 4

U_i	Forecast of the ith ensemble member, for wind component U
V_i	Forecast of the ith ensemble member, for wind component V
$\overline{U}$	Ensemble mean for wind component U
$\overline{V}$	Ensemble mean for wind component V
S_U/S_U^2	Ensemble standard deviation and variance for wind component U
S_V/S_V^2	Ensemble standard deviation and variance for wind component V
$\mathbf{x}$	Observed wind vector
U	Observed value for wind component U
V	Observed value for wind component V
$\boldsymbol{\mu}$	Predictive mean vector
μ_U	Predictive mean for wind component U
μ_V	Predictive mean for wind component V
$\boldsymbol{\Sigma}$	Predictive variance-covariance matrix
σ_U/σ_U^2	Predictive standard deviation and variance for wind component U

σ_V/σ_V^2	Predictive standard deviation and variance for wind component V
ρ	Predictive correlation coefficient of the wind components
θ	Predictive wind direction, computed from the predictive mean
w	Observed wind speed
M	Ensemble size
m	Length of training period in days
s	Location for which the forecasts is issued
t	Valid time of the forecast

Bibliography

Anderson, J. L. (1996) A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, **9**, 1518–1530.

Baars, J. (2005) Observations QC documentation. Available at `http://www.atmos.washington.edu/mm5rt/qc_obs/qc_doc.html`.

Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer and T. Reinhardt (2011) Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities. *To appear in Monthly Weather Review*.

Bao, L., T. Gneiting, E. P. Grimit, P. Guttorp and A. E. Raftery (2010) Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction. *Monthly Weather Review*, **138**, 1811–1821.

Bernardo, J. M. (1979) Expected information as expected utility. *Annals of Statistics*, **7**, 686–690.

Berrocal, V. J., A. E. Raftery and T. Gneiting (2007) Combining spatial statistical and ensemble information in probabilistic weather forecasts. *Monthly Weather Review*, **135**, 1386–1402.

Bröcker, J. and L. A. Smith (2007) Scoring probabilistic forecasts: The importance of being proper. *Weather and Forecasting*, **22**, 382–388.

Dawid, A. (1984) Statistical theory: The prequential approach. *Journal of the Royal Statistical Society Series A*, **147**, 278–292.

Delle Monache, L., J. P. Hacker, Y. Zhou, X. Deng and R. B. Stull (2006) Probabilistic aspects of meteorological and ozone regional ensemble forecasts. *Journal of Geophysical Research*, **111**, D24307, doi:10.1029/2005JD006917.

Diebold, F. X., T. A. Gunther and A. S. Tay (1998) Evaluating density forecasts with applications to financial risk management. *International Economic Review*, **39**, 863–883.

Eckel, F. A. and C. F. Mass (2005) Aspects of effective mesoscale, short-range ensemble forecasting. *Weather and Forecasting*, **20**, 328–350.

Gigerenzer, G., R. Hertwig, E. van den Broek, B. Fasolo and K. V. Katsikopoulos (2005) "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis*, **25**, 623–9.

Glahn, H. R. and D. A. Lowry (1972) The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology*, **11**, 1203–1211.

Gneiting, T. (2008) Editorial: Probabilistic forecasting. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **171**, 319–321.

Gneiting, T. (2011) Making and evaluating point forecasts. *Journal of the American Statistical Association*, **106**, 746–762.

Gneiting, T., F. Balabdaoui and A. E. Raftery (2007) Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69**, 243–268.

Gneiting, T. and A. E. Raftery (2005) Weather forecasting with ensemble methods. *Science*, **310**, 248.

Gneiting, T. and A. E. Raftery (2007) Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359–378.

Gneiting, T., A. E. Raftery, A. H. Westveld and T. Goldman (2005) Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, **133**, 1098–1118.

Gneiting, T., L. I. Stanberry, E. P. Grimit, L. Held and N. A. Johnson (2008) Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test*, **17**, 211–235.

Good, I. (1952) Rational decisions. *Journal of the Royal Statistical Society Series B*, **14**, 107–114.

Grimit, E. and C. Mass (2002) Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Weather and Forecasting*, **17**, 192–205.

Grimit, E. P. (2001) *Implementation and Evaluation of a Mesoscale Short-Range Ensemble Forecasting System over the Pacific Northwest*. Master's thesis, Department of Atmospheric Sciences, University of Washington.

Hamill, T. M. (2001) Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, **129**, 550–560.

Hamill, T. M. and S. J. Colucci (1997) Verification of Eta-RSM short-range ensemble forecasts. *Monthly Weather Review*, **125**, 1312–1327.

Jolliffe, I. T. (2008) Comments on: Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test*, **17**, 249–250.

Leutbecher, M. and T. Palmer (2008) Ensemble forecasting. *Journal of Computational Physics*, **227**, 3515–3539.

Matheson, J. E. and R. L. Winkler (1976) Scoring rules for continuous probability distributions. *Management Science*, **22**, 1087–1096.

National Weather Service (1998) Automated Surface Observing System (ASOS) User's Guide. Available at `http://www.weather.gov/asos/aum-toc.pdf`.

Palmer, T. (2002) The economic value of ensemble forecasts as a tool for risk assessment: from days to decades. *Quarterly Journal of the Royal Meteorological Society*, **128**, 747–774.

Pinson, P. (2011) Adaptive calibration of (u ,v)-wind ensemble forecasts. Available at `http://www2.imm.dtu.dk/~pp/docs/pinson11_uvcalibrev.pdf`.

R Development Core Team (2011) R: A Language and Environment for Statistical Computing. Available at `http://www.r-project.org`.

Raftery, A. E., T. Gneiting, F. Balabdaoui and M. Polakowski (2005) Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, **133**, 1155–1174.

Roulston, M. and L. Smith (2003) Combining dynamical and statistical ensembles. *Tellus A*, **55**, 16–30.

Savage, L. J. (1971) Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, **66**, 783–801.

Schefzik, R. (2011) *Ensemble Copula Coupling*. Diploma thesis, Faculty of Mathematics and Computer Science, Heidelberg University.

Sloughter, J. M. (2009) Probabilistic weather forecasting using Bayesian model averaging. PhD final presentation, available at `http://www.atmos.washington.edu/~cliff/FinalExamination-20090508-JMcLeanSloughter.ppt`.

Sloughter, J. M., T. Gneiting and A. E. Raftery (2010) Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *Journal of the American Statistical Association*, **105**, 25–35.

Sloughter, J. M., A. E. Raftery, T. Gneiting and C. Fraley (2007) Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Monthly Weather Review*, **135**, 3209–3220.

Talagrand, O., R. Vautard and B. Strauss (1997) Evaluation of probabilistic prediction systems. *Proc. Workshop on Predictability*, 1–25, Reading, UK, European Centre for Medium-Range Weather Forecasts.

Thorarinsdottir, T. L. and T. Gneiting (2010) Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroscedastic censored regression. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **173**, 371–388.

Thorarinsdottir, T. L. and M. S. Johnson (2011) Probabilistic wind gust forecasting using non-homogeneous Gaussian regression. *To appear in Monthly Weather Review*.

Vardi, Y. and C.-H. Zhang (2000) The multivariate L1-median and associated data depth. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 1423–6.

Whitaker, J. S. and A. F. Loughe (1998) The relationship between ensemble spread and ensemble mean skill. *Monthly Weather Review*, **126**, 3292–3302.

Wilks, D. S. (1995) *Statistical Methods in the Atmospheric Sciences*. Academic Press.

Winkler, R. L. (1977) Rewarding expertise in probability assessment. *Decision Making and Change in Human Affairs* (eds. H. Jungermann and G. de Zeeuw), 127–140, Dordrecht: Reidel.

Winkler, R. L. (1996) Scoring rules and the evaluation of probabilities. *Test*, **5**, 1–60.

Danksagung

Ich möchte Dr. Thordis L. Thorarinsdottir und Prof. Dr. Tilmann Gneiting dafür danken, dass sie es mir ermöglicht haben mich mit diesem interessanten Thema zu beschäftigen und dass sie mich, nicht nur im Zuge dieser Arbeit, so hervorragend betreut haben. Ein ganz besonderer Dank gilt meiner Familie, die mich immer uneingeschränkt unterstützt hat und mir mit Rat und Tat zur Seite stand. Zuletzt möchte ich mich noch herzlich bei Kira Feldmann und Corinna Frei für die tolle und produktive Zusammenarbeit bedanken.

Erklärung

Hiermit versichere ich, dass ich meine Arbeit selbstständig unter Anleitung verfasst habe, dass ich keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, und dass ich alle Stellen, die dem Wortlaut oder dem Sinne nach anderen Werken entlehnt sind, durch die Angabe der Quellen als Entlehnungen kenntlich gemacht habe.

15. September 2011