NIH Public Access
Author Manuscript
*Environmetrics*. Author manuscript; available in PMC 2013 December 16.

Published in final edited form as:
*Environmetrics*. 2012 February 1; 23(1): . doi:10.1002/env.1133.

# Circular Conditional Autoregressive Modeling of Vector Fields*

**Danny Modlin**,
Statistics PhD student at North Carolina State University (NCSU)

**Montse Fuentes**, and
Professor of Statistics at NCSU. Tel: (919) 515-1921, Fax: (919) 515-1169

**Brian Reich**
Assistant Professor of Statistics at NCSU

Danny Modlin: drmod-lin@ncsu.edu; Montse Fuentes: fuentes@stat.ncsu.edu

## Abstract

As hurricanes approach landfall, there are several hazards for which coastal populations must be prepared. Damaging winds, torrential rains, and tornadoes play havoc with both the coast and inland areas; but, the biggest seaside menace to life and property is the storm surge. Wind fields are used as the primary forcing for the numerical forecasts of the coastal ocean response to hurricane force winds, such as the height of the storm surge and the degree of coastal flooding. Unfortunately, developments in deterministic modeling of these forcings have been hindered by computational expenses. In this paper, we present a multivariate spatial model for vector fields, that we apply to hurricane winds. We parameterize the wind vector at each site in polar coordinates and specify a circular conditional autoregressive (CCAR) model for the vector direction, and a spatial CAR model for speed. We apply our framework for vector fields to hurricane surface wind fields for Hurricane Floyd of 1999 and compare our CCAR model to prior methods that decompose wind speed and direction into its N-S and W-E cardinal components.

### Keywords

CAR; circular statistics; cross-covariance; hurricane winds; spatial statistics

## 1 Introduction

Across many areas of research, one may come into contact with vector data. One such type of data is wind fields. Studying wind fields is important in environmental research. For example, with the current insurgence of support for cleaner energy, there is an increased focus on studying spatial and temporal variations in wind speed and direction (Hering and Genton, 2010). Researchers are trying to identify optimal locations for wind turbines, which requires a model for the wind speed, direction, and duration at different sites. Another example is the wind fields generated by a hurricane. Residents living along the coastal area of the southeast United States and Gulf Coast are presented with many hazards during a landfalling hurricane. With populations in these areas increasing, it is imperative that as storms approach the coastline we have the means necessary to give these citizens the information needed to prepare for possible landfall conditions. Storm winds, torrential rain, and spawned tornadoes each can harm both life and property, but the single largest threat to

*The authors thank the National Science Foundation (Fuentes DMS-0706731, CMG-0934595), the Environmental Protection Agency (Fuentes, R833863), and National Institutes of Health (Fuentes, 5R01ES014843-02) for partial support of this work. The authors would also like to thank Sujit Ghosh, Professor of Statistics at NCSU, for his input and conversation.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

coastal areas is the storm surge. This inundation of water pushed by the landfalling storm can quickly take lives and destroy property with homes and businesses being either right at or just a few feet above sea level.

Developing accurate forecasts of storm surges will improve the preparedness of these communities. Research has shown that the effectiveness of these forecasts depends upon accurate modeling of wind forcings. At present, there has not been a model adopted for the forecasting/mapping hurricane winds for the specific purpose of improving storm surge forecasts. Some researchers believe that this is due to the computational expense of such models. However, there have been some models and methods developed to assist with modeling these hurricane wind fields. Holland (1980), Depperman (1947), and DeMaria et al. (1992) each presented models that have been termed as axis-symmetric. These models are based upon a cyclostrophic wind balance and place the key dependence on the distance a location is from the storm circulation center. These models are simple to understand and apply; however, they do not describe the true asymmetrical structure of the winds within hurricanes. For example, winds in the northeast quadrant of the storm are typically stronger than those in other locations due to friction, environment, vertical shear, etc. These and other possible sources were discussed by Chen and Yau (2003), Ross and Kurihara (1992), Shapiro (1983), and Wang and Holland (1980).

Xie et al. (2011) looked into the effect asymmetry of a storm has on the storm surge. Using the Coastal Marine Environmental Prediction System (CMEPS), developed at North Carolina State University, they simulated storm surge under different conditions. They found that there was significant difference in water levels when the asymmetry of hurricane wind fields was changed while holding all other parameters such as maximum wind speed, radius of maximum winds, and minimum pressure constant. Xie et al. commented that there has been improvements made to storm surge forecasts. However, they point out that all of these advancements were made under the assumption that the wind forcing fields were accurate.

With the knowledge that hurricanes are asymmetrical, there have been models proposed that would attempt to incorporate asymmetric structures (Georgiou, 1985). Xie et al. (2006) looked at the wind model developed by Holland and attempted to model its error utilizing a Gaussian process. Reich and Fuentes (2007) took this approach one step further and removed the assumption of the Gaussian process. With the incorporation of a stick-breaking prior, they were able to develop a non-parametric approach that was more general than the previous. This model assumed that the cross-dependence between the N-S and W-E wind components was constant across space which may not hold in practice.

The common statistical modeling approach for wind vectors decomposes the wind fields into the $u$ and $v$ components (Cartesian representation), where $u$ corresponds to the N-S and $v$ to the W-E wind component. Figure 1 plots the data for Hurricane Floyd on September 14, 1999. Modeling $u$ (Figure 2a) is challenging because it displays heavy-tails (non-normality), increased variability, and a shorter spatial range (non-stationarity) near the storm center. Joint modeling of $u$ and $v$, as in Reich and Fuentes (2007), is also complicated because their correlation varies dramatically in different parts of the spatial domain. In contrast, the logarithm transformation of wind speed and wind direction, respectively, vary relatively smoothly in space and do not have a complicated joint relationship (Figure 3). Therefore, in this paper we propose to model hurricane wind fields using polar coordinates.

Modeling wind using polar coordinates presents challenges of its own. The literature on spatial modeling of angles is limited. Morphet (2009) presents some frequentist methods as well as enhanced the visualization of circular-spatial data through the development of an R

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

package. Morphet developed a circular kriging solution that was based on fitting a new defined cosineogram. Morphet also presented a method of simulating from a circular random field that was a transformation of a Gaussian random field.

The Bayesian approach for hurricane modeling has several advantages, including a convenient framework for simultaneously modeling several data sources (e.g., satellite and buoy data) and natural measures of uncertainty for model parameters, which are crucial inputs to deterministic hurricane and storm surge models. Ravindran (2002) approaches circular data from a Bayesian perspective utilizing wrapped distributions. Ravindran states that likelihood-based inference for these wrapped distributions can be very complicated and not be computationally efficient. These issues are resolved using Markov Chain Monte Carlo (MCMC) method with a data augmentation step. An extension is given for time-correlated data. To our knowledge, we present the first hierarchical Bayesian model for spatial circular data.

In this paper, we present a new statistical modelling framework for spatial vector fields, for hurricane wind fields. With the assistance of wrapped distributions, we model the angle of the wind direction using a circular conditional autoregressive model (CCAR). The wind speed and wind direction at a particular location within a storm tend to be less correlated than the *u* and *v* components. Then, it is easier to explain the spatial cross-dependence of wind vectors using polar coordinates. The paper is organized as follows. In Section 2 we review circular statistics. In Section 3 we describe the new CCAR methodology. In Section 4 we apply our methods to Hurricane Floyd. We conclude with results and some final remarks in Section 5 and 6 respectively.

## 2 Circular Statistics

Since the 1970s, there have been advancements in the analysis of circular data with a "vigorous development" of methods in the 1980s (Fisher, 1993). Angles are vastly different than their linear counterparts. Computation of summary statistics, performing analysis, and simply displaying the data all must take into account their periodic nature. Thus the standard approaches to model distributions and calculate moments have to be modified when working with angles. This section describes the common approaches to obtain moments and distributions of angles; for further information, we refer to Fisher (1993).

### 2.1 Sample Moments

We begin with the calculation of the mean. With linear data, the sample mean is $\overline{x}=\sum_{i=1}^{n}\frac{x_i}{n}$. When $x_i = \theta_i$ is an angle, this is not appropriate because this ignores similarity of values near 0 and near $2\pi$. For the angular mean, it is more appropriate to use vector addition. We begin by calculating three values: $C=\sum_{i=1}^{n}\cos\theta_i$, $S=\sum_{i=1}^{n}\sin\theta_i$, and $R^2 = C^2 + S^2$. With these calculations, the value (direction) of $\theta$ is $\overline{\theta}=\arctan\frac{S}{C}$.

$R=\sqrt{C^2+S^2}\in(0,n)$ is commonly referred to as the resultant length of the resultant vector. Thus, we can calculate the mean resultant length $\overline{R}=\frac{R}{n}$. $\bar{R} = 1$ represents all the points were overlapping; however, $\bar{R} = 0$ does not imply uniform dispersion around the circle. The main usage of $\bar{R}$ is in the calculation of sample circular variance, $V = 1 - \bar{R}$. Similar to the interpretation of linear variance, a small circular variance does imply that the distribution of data was more concentrated. Differences between linear and circular variance are that $V \in [0, 1]$ and calculation of standard deviation is not just a square root. Sample circular standard deviation is defined as $\{-2\log(1-V)\}^{\frac{1}{2}}$. These calculations are needed for calculating posterior means and standard deviations from MCMC output.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

### 2.2 Circular Distributions

Many distributions can be placed on circular data. The common way to generate a circular distribution is wrapping distributions on $\Re$ to the unit circle. If $X$ is a random variable on the real line, we can construct a random variable on the circle and determine its density. Assume that $X$ has probability density function $g(x)$ and cumulative distribution function $G(x)$, and define $\theta = X[mod 2\pi]$. The probability density function of $\theta$, $f(\theta)$, is found by wrapping $g(x)$ around a unit circle. Thus $f(\theta)=\sum_{k=-\infty}^{\infty} g(\theta+2k\pi)$ with corresponding distribution of $F(\theta)=\sum_{k=-\infty}^{\infty}[G(\theta+2k\pi)-G(2k\pi)]$. The Wrapped Normal distribution with $X \sim N(\mu, \sigma^2)$ is of particular interest in modeling wind fields. The density is

$$f(\theta)=\sum_{j=-\infty}^{\infty} \varphi(\theta|2\pi j+\mu,\sigma^2), \quad (1)$$

where $\varphi(\cdot|m, s^2)$ is the $N(m, s^2)$ density function. In the wrapped normal model, the mean direction is $E(\theta_i) = \mu_i$ and $\sigma^2 > 0$ controls the variability. We denote this model as $\theta \sim WN(\mu, \sigma^2)$.

## 3 Hierarchical Bayesian spatial model for a vector field

We assume that the response in grid cell $i = 1, \ldots, n$ is a vector defined by its speed $\omega_i$ and direction $\theta_i$. Therefore we model $y_i = \log(\omega_i) \in \mathcal{R}$. Our circular conditional autoregressive model (CCAR) statistical framework is as follows:

$$\begin{aligned} y_i|\theta_i &\sim N(\mathbf{X}_i^T\beta_1+g(\theta_i)+\mu_{1i},\sigma_1^2) \\ \theta_i &\sim WN(\mathbf{X}_i^T\beta_2+\mu_{2i},\sigma_2^2) \end{aligned} \quad (2)$$

where $\mathbf{X}_i^T\beta_1$ and $\mathbf{X}_i^T\beta_2$ represent the contribution to each mean by covariates, $g(\theta_i)$ captures the relationship between speed and direction after accounting for covariates, and $\mu_{i1}$ and $\mu_{2i}$ are spatial effects. $\mathbf{X}_i$ includes covariates radial distance from center of storm $r_i$, latitude of location $i$, longitude of location $i$, and the sine and the cosine of the inflow angle at cell $i$ across circular isobars towards the storm center $\phi_i \in [0, 2\pi)$. Currently no covariates are included for the variability of the errors or spatial effect for angular data. Our model can be adapted to include covariates to explain the variability of wind speed and direction. For example, the log of the variance parameter could be modeled with a linear relationship with radial distance from center.

There are several possibilities for a functional relation between the $\theta_i$ and $y_i$ that could be included in $g(\theta)$. One could assume a linear mean model $g(\theta_i) = b\theta_i$, but this is not appropriate, since conceptually we should have $g(0) = g(2\pi)$. Another is the standard approach for circular/linear association is $g(\theta_i) = b\cos(\theta_i)$ (Fisher, 1993). To specify more complicated circular/linear relationship, rather than including higher-order polynomials, one could include higher frequencies $g(\theta_i)=\sum_{k=1}^{M} a_k\sin(k\theta_i)+\sum_{k=1}^{M} b_k\cos(k\theta_i)$.

Modeling $\theta_i$ is challenging due to the restrictions that $\theta_i \in [0, 2\pi)$ and that its density at 0 and $2\pi$ should be equal since these are the same angle. We model $\theta_i$ by extending the wrapped normal (WN) distribution to the spatial setting. The WN distribution alleviates several difficulties in modeling spatially-referenced angles. Unfortunately, the WN density (1) cannot be evaluated directly because it includes an infinite sum with no closed form. However, we are able to analyze this model using MCMC methods after introducing

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

auxiliary variables for the wrap number, $K_i \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. To simplify notation, define $\overline{\mu}_i = \mathbf{X}_i^T \beta_2 + \mu_{2i}$. The auxiliary model is

$$\begin{aligned} \theta_i | K_i \sim \mathrm{TN}_{[0,2\pi]} \left(2\pi K_i + \overline{\mu}_i, \sigma^2\right) \\ P(K_i = j) = \Phi(2\pi | 2\pi j + \overline{\mu}_i, \sigma^2) - \Phi(0 | 2\pi j + \overline{\mu}_i, \sigma^2), \end{aligned} \quad (3)$$

where $\mathrm{TN}_A(m, s^2)$ denotes the truncated normal distribution with domain $A$, location $m$, and scale $s$, and $\Phi(\cdot | m, s^2)$ is the distribution function of a normal with mean $m$ and standard deviation $s$. The truncated normal density can be written

$$p(\theta_i | K_i = j) = \frac{\varphi(\theta_i | 2\pi j + \overline{\mu}_i, \sigma^2)}{\Phi(2\pi | 2\pi j + \overline{\mu}_i, \sigma^2) - \Phi(0 | 2\pi j + \overline{\mu}_i, \sigma^2)} = \frac{\varphi(\theta_i | 2\pi j + \overline{\mu}_i, \sigma^2)}{P(K_i = j)}. \quad (4)$$

Therefore, marginally over $K_i$,

$$p(\theta_i) = \sum_{j=-\infty}^{\infty} \frac{\varphi(\theta | 2\pi j + \overline{\mu}_i, \sigma^2)}{P(K_i = j)} P(K_i = j) = \sum_{j=-\infty}^{\infty} \varphi(\theta_i | 2\pi j + \overline{\mu}_i, \sigma^2), \quad (5)$$

as desired.

Clearly $\sum_{j=-\infty}^{\infty} P(K_i = j) = 1$. However, implementing this prior is challenging since $K_i$ has an infinite domain and non-standard prior. An equivalent representation of (3) is

$$\begin{aligned} \theta_i | z_i \sim \mathrm{TN}_{[0,2\pi]} \left(2\pi \left(\lfloor -z_i/(2\pi) \rfloor + 1\right) + \overline{\mu}_i, \sigma^2\right) \\ z_i \sim N(\overline{\mu}_i, \sigma^2), \end{aligned} \quad (6)$$

where $\lfloor -z_i/(2\pi) \rfloor$ is defined as the largest integer less than $-z_i/(2\pi)$. Here we replace $K_i$'s prior in (3) with the two-stage model $K_i = \lfloor -z_i/(2\pi) \rfloor + 1$ and $z_i \sim N(\overline{\mu}_i, \sigma^2)$, which gives the same prior probabilities for $K_i$ since

$$\begin{aligned} P(K_i = j) &= P(\lfloor -z_i/(2\pi) \rfloor + 1 = j) \\ &= P(j - 1 < -z_i/(2\pi) < j) \\ &= P(-2\pi j < z_i < -2\pi(j-1)) \\ &= \Phi(-2\pi(j-1) | \overline{\mu}_i, \sigma^2) - \Phi(-2\pi j | \overline{\mu}_i, \sigma^2) \\ &= \Phi(2\pi | 2\pi j + \overline{\mu}_i, \sigma^2) - \Phi(0 | 2\pi j + \overline{\mu}_i, \sigma^2). \end{aligned} \quad (7)$$

As shown in the Appendix, this representation is conducive to standard software packages because it only requires standard parametric distributions.

The simplest setup for spatial random effects $\mu_{1i}$ and $\mu_{2i}$ is a proper conditionally autoregressive prior ("CAR"; Banerjee et al., 2004). The CAR covariance is specified through spatial adjacencies. Let $i \sim j$ indicate that cells $i$ and $j$ are spatial neighbors and $m_i$ be the number of spatial neighbors of cell $i$. The CAR model for the log vector lengths $\mu_{1i}$ is defined through the full conditional distribution of $\mu_{1i}$ given $\mu_{1j}$ at all other cells with $j \neq i$. The full conditional distribution is Gaussian with

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$E[\mu_{1i}|\mu_{1j}, j \neq i] = \rho_1 \sum_{j \sim i} \mu_{1j}/m_i$$
$$V[\mu_{1i}|\mu_{1j}, j \neq i] = \tau_1^2/m_i. \quad (8)$$

The full conditional mean is proportional to the average of the spatial neighbors, $\rho_1 \in [0, 1]$ controls the degree of spatial association, and the variance is controlled by $\tau_1^2 > 0$. We denote the model $\mu_1 = (\mu_{11}, \ldots, \mu_{1n})^T \sim \mathrm{CAR}(\rho_1, \tau_1^2)$.

A potential prior for the spatial angle effects are $\mu_2 = (\mu_{21}, \ldots, \mu_{2n})^T \sim \mathrm{CAR}(\rho_2, \tau_2^2)$. However, using this direct CAR modeling of the angular spatial random effect $\mu_{2i}$ may be problematic. The central assumption of the CAR model is that the full conditional distributions are centered on the linear mean of neighboring regions. For angular data, the conditional mean in (8) may not be a good summary of the neighboring angles. For example, if half of the neighbors are slightly above 0 and half are slightly less than $2\pi$, the linear mean in (8) is $\pi$, when in fact the mean should be near zero.

Therefore, we need an alternative manner of modeling the spatial random effect $\mu_{2i}$ that respects the periodicity of the circular spatial process. To define the angle model, we introduce two latent spatial processes $S_i$ and $C_i$, each with proper CAR priors $S = (S_1, \ldots, S_n) \sim \mathrm{CAR}(\rho_2, \tau_2^2)$ and $C = (C_1, \ldots, C_n) \sim \mathrm{CAR}(\rho_3, \tau_3^2)$. From each of these, we perform a hyperbolic tangent transformation,

$$\overline{S}_i = 2\frac{\exp(S_i)}{1+\exp(S_i)} - 1 \quad \text{and } \overline{C}_i = 2\frac{\exp(C_i)}{1+\exp(C_i)} - 1, \quad (9)$$

so that both $\overline{S}_i$ and $\overline{C}_i$ are on [−1, 1]. Here $\overline{S}_i$ and $\overline{C}_i$ represent the sine and cosine, respectively, of the angle process. The value of $\mu_{2i}$ is then calculated using the inverse tangent,

$$\mu_{2i} = \arctan\left(\frac{\overline{S}_i}{\overline{C}^i}\right) \quad (10)$$

To ensure that $\mu_{2i} \in [0, 2\pi)$, we adjust the previous result by adding $\pi$ when $\overline{C}_i < 0$ and adding $2\pi$ when $\overline{S}_i < 0$ and $\overline{C}_i > 0$.

Modeling the sine and cosine of $\mu_{2i}$ alleviates the problem with the usual CAR described above. For example, in the scenario with the neighboring observations split between small positive values and values slightly below $2\pi$, $\overline{S}_i$ will be near zero for all neighbors and $\overline{C}_i$ will be near one for all neighbors. Therefore the full conditional priors for $\overline{S}_i$ and $\overline{C}_i$ will be near zero and one, respectively, correctly centering the prior for $\mu_{2i} = \arctan\left(\frac{\overline{S}_i}{\overline{C}_i}\right)$ on zero.

To complete the Bayesian model, we specify uninformative priors for the hyperparameters. We use independent N(0, 100) priors for the elements of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, independent InvGamma(0.5,0.0005) prior of the variances $\sigma_1^2, \sigma_2^2, \tau_1^2, \tau_2^2$, and $\tau_3^2$ and independent Unif(0,1) priors of the CAR association parameters $\rho_1$, $\rho_2$, and $\rho_3$. MCMC sampling is performed using WinBUGS. We run chains of length 20,000, and discard the first 5,000 samples as burn-in. Convergence is monitored using trace plots and autocorrelations for the deviance and several representative parameters.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

## 4 Analysis of Hurricane Floyd

We use satellite data obtained from NOAA to characterize wind fields. The data are publicly available at www.ncdc.noaa.gov/oa/rsad/seawinds.html. This is the best source of hurricane wind satellite data that we currently have. It is obtained by combining different satellites, and it is stored across the globe on a grid of 0.25 degree squares. We will focus on the September 14th noon observance of Hurricane Floyd, a category three storm, from the 1999 hurricane season. Our area of interest is a 41×41 grid centered approximately on the storm's center of circulation.

Our data is given in the Cartesian decomposition format therefore we transform to the polar scale, $\omega_i = \sqrt{u_i^2 + v_i^2} > 0$ and $\theta_i = \arctan \frac{u_i}{v_i} \in [0, 2\pi)$. For our CCAR model, recall that $y_i$ = log $\omega_i$. Empirical analysis indicates that a log transform of the vector speed allowed conditional normality to be a reasonable assumption. Figure 4 shows a qq-plot of the residuals from a model under the log transformation of wind speed. As we can see, these residuals closely approximate a straight line. Figure 5 shows a scatterplot of wind direction against the residuals of the log transform of wind speed after accounting for $\mathbf{X}_i$. No discernable pattern can be seen, therefore the covariates mentioned above appear to capture the effects of direction on speed and we take $g(\theta_i) = 0$.

We compare the following two models:

1. **Circular model**: $y_i | \theta_i \sim N(\mathbf{X}_i^T \beta_1 + \mu_{1i}, \sigma_1^2)$ and $\theta_i \sim \mathrm{WN}(\mathbf{X}_i^T \beta_2 + \mu_{2i}, \sigma_2^2)$
2. **U/V model**: $u_i \sim N(\mathbf{X}_i^T \beta_3 + \mu_{ui}, \sigma_u^2)$ and $v_i \sim N(\mathbf{X}_i^T \beta_4 + a\mu_{ui} + \mu_{vi}, \sigma_v^2)$

where $\mathbf{X}_i$ includes covariates such as radial distance from center of storm $r_i$, latitude of location $i$, longitude of location $i$, and the sine and the cosine of the inflow angle at cell $i$ across circular isobars towards the storm center $\phi_i \in [0, 2\pi)$.

## 5 Results

We compare the performance of our CCAR model to the standard U/V model through five-fold cross validation. Our original dataset is partitioned into five randomly selected groups. Each group, in turn, serves as the validation dataset with the other four serving as the training set. The mean of the posterior realizations is calculated at the missing points and compared with the observed values using the metrics explained below. With the category of the storm dependent on the magnitude of the fastest wind vector, our focus is in the calculation of the wind speed and direction. For each of these models the posterior mean of $\hat{\omega}$ is calculated and then summarized using the mean square error (MSE). The posterior mean of direction, $\hat{\theta}$, is calculated using the methods described in Section 2. $\hat{\theta}$ is compared using two metrics. First, we use the mean absolute cosine error (MACE),

$$\mathrm{MACE} = \frac{1}{n} \sum_{i=1}^{n} |\cos(\theta_i) - \cos(\hat{\theta}_i)| \quad (11)$$

where closer to 0 indicates a better model. The other metric is mean cosine difference error (MCDE),

$$\mathrm{MCDE} = \frac{1}{n} \sum_{i=1}^{n} \cos(\theta_i - \hat{\theta}_i) \quad (12)$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

in this case closer to 1 is better. Table 1 gives the calculated MSE, MACE, and MCDE values for the U/V model and the CCAR model along with their standard errors. We see that there is pronounced significant improvement, of almost 30 times, in the modeling of the wind speed. When we compare the direction, we see that the CCAR models outperforms the U/V model in both MACE and MCDE.

We compared the observed direction with that of the posterior mean calculated within our five-fold cross validation. Figure 6 shows an arrow plot of the observed directions, solid arrow, compared to the posterior mean direction, dashed arrow. The worst mistakes are made on points that are closer to the eye of the storm. For all other points within the storm, our model performs well.

### 5.1 Sensitivity Analysis

A separate sensitivity analysis was performed to determine the effect of the selection of the prior for the variance components on the results. Gelman (2006) discussed the possible prior distributions that could be utilized for scale parameters. We consider three priors for the scale parameters of our model, Gamma(0.5,0.0005), Uniform(0,100), and Uniform(0,300). We compare these priors under the same MACE and MCDE as described above. Table 2 displays the results of our sensitivity analysis; prediction errors are not significantly different for the three priors.

## 6 Discussion and Remarks

In this paper, we present an innovative multivariate fully-Bayesian spatial model for vector fields, that we apply to hurricane winds. We introduce for the first time in the literature a spatial version of circular distributions, a circular conditional autoregressive (CCAR) model for the vector direction utilizing wrapped distributions. We implemented our framework for vector fields to better characterize hurricane surface wind fields. A case study of Hurricane Floyd of 1999 showed that our CCAR model outperformed prior methods that decompose wind speed and direction into its N-S and W-E cardinal components.

We analyze, in our case study, only responses from a single source, blended satellite data. A second source of data, buoy measurements, can also be included to be combined with the satellite data. In our CAR model, we utilized the standard proximity neighborhood structure. As future work we can introduce a neighborhood structure that would be more representative of the true neighbors within hurricane wind fields, perhaps define neighbors based on polar coordinates of the grid cells relative to the storm center. Our case study analyzed only one time point of Hurricane Floyd's track towards the US East Coast. Our model could be altered to account for time series data while still accounting for the spatial structure across our region of the Atlantic Basin.

## References

Banerjee, S.; Carlin, BP.; Gelfand, AE. Hierarchical modeling and analysis for spatial data. Chapman and Hall; New York: 2004.

Chen Y, Yau MK. Asymmetric structures in a simulated landfalling hurricane. Journal of Atmospheric Sciences. 2003; 60:2294–2312.

DeMaria M, Aberson SD, Ooyama KV, Lord SJ. A nested spectral model for hurricane track forecasting. Monthly Weather Review. 1992; 120:1628–1640.

Depperman RC. Notes on the origin and structures of Philippine typhoons. Bulletin of the American Meteorological Society. 1947; 28:399–404.

Fisher, NI. Statistical Analysis of Circular Data. Cambridge University Press; Cambridge: 1993.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Gelman A. Prior Distributions for Variance Parameters in Hierarchical Models. Bayesian Analysis. 2006; 1(3):515–533.

Georgiou, P. PhD thesis. University of Western Ontario; 1985. Design Wind Speeds in Tropical Cyclone Prone Regions.

Hastings, C, Jr. Approximations for digital computers. Princeton University Press; Princeton: 1955.

Hering A.S. and Genton M.G. Powering up with space-time wind forecasting. Journal of the American Statistical Association. 2010; 105:92–104.

Holland G.J. An analytic model of the wind and pressure profiles in hurricanes. Monthly Weather Review. 1980; 108:1212–1218.

Morphet, WJ. PhD thesis. Utah State University; 2009. Simulation, Kriging, and Visualization of Circular-Spatial Data.

Ravindran, P. PhD thesis. North Carolina State University; 2002. Bayesian Analysis of Circular Data Using Wrapped Distributions.

Reich B. and Fuentes M. A Multivariate Semiparametric Bayesian Spatial Modeling Framework for Hurricane Surface Wind Fields. The Annals of Applied Statistics. 2007; 1:249–264.

Ross R.J. and Kurihara Y. A simplified scheme to simulate asymmetries due to the beta effect in barotropic vortices. Journal of Atmospheric Sciences. 1992; 49:1620–1628.

Shapiro L. The asymmetric boundary layer flow under a translating hurricane. Journal of Atmospheric Sciences. 1983; 40:1984–1998.

Wang Y. and Holland G.L. Tropical cyclone motion and evolution in vertical shear. Journal of Atmospheric Sciences. 1996; 53:3313–3332.

Xie L, Bao S, Pietrafesa L, Foley K, Fuentes M. A real-time hurricane surface wind forecasting model: Formulation and verification. Monthly Weather Review. 2006; 134:1355–1370.

Xie L, Liu H, Liu B, Bao S. A numerical study of the effect of hurricane wind asymmetry on storm surge and inundation. Ocean Modeling. 2011; 36:71–79.

## Appendix - The wrapped-normal density in WinBUGS

In this section we give the WinBUGS code used to specify the WN density for the vector angles. The remaining code for the vector length model, and all hyperpriors are omitted since they are straight-forward to code in WinBUGS. The WN likelihood is given by

```
for(i in 1:n){
y[i]~dnorm(meany[i],tau2)
one[i]<-1
one[i]~dbern(denom[i])
meany[i]<-mu2[i] + 2*pi*K[i]
K[i]<-trunc(-z[i]/(2*pi))+1
z[i]~dnorm(mu2[i],tau2)
U[i]<-phi(sqrt(tau)*(2*pi-mu2[i]))
L[i]<-phi(sqrt(tau)*(0-mu2[i]))
denom[i]<-c/(U[i]-L[i])
}
```

where $c = \exp(-200)$ is a small constant used in the “ones trick” to ensure that denom[i] $\in (0, 1)$. The product of the normal density for y[i] and the Bernoulli density for one[i] gives the truncated normal density in (4). The combination of models for K[i] and z[i] executes the auxiliary model in (6).

Both the U/V model and our CCAR model were implemented in WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/). Due to inverse trigonometric functions not being available

in WinBUGS, we use an inverse sine approximation designed by C. Hastings, Jr. (1955) on $|\sin[i]|$. This polynomial form allowed us to remain in WinBUGS for the entirety of the execution, thus maintaining computational convenient.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

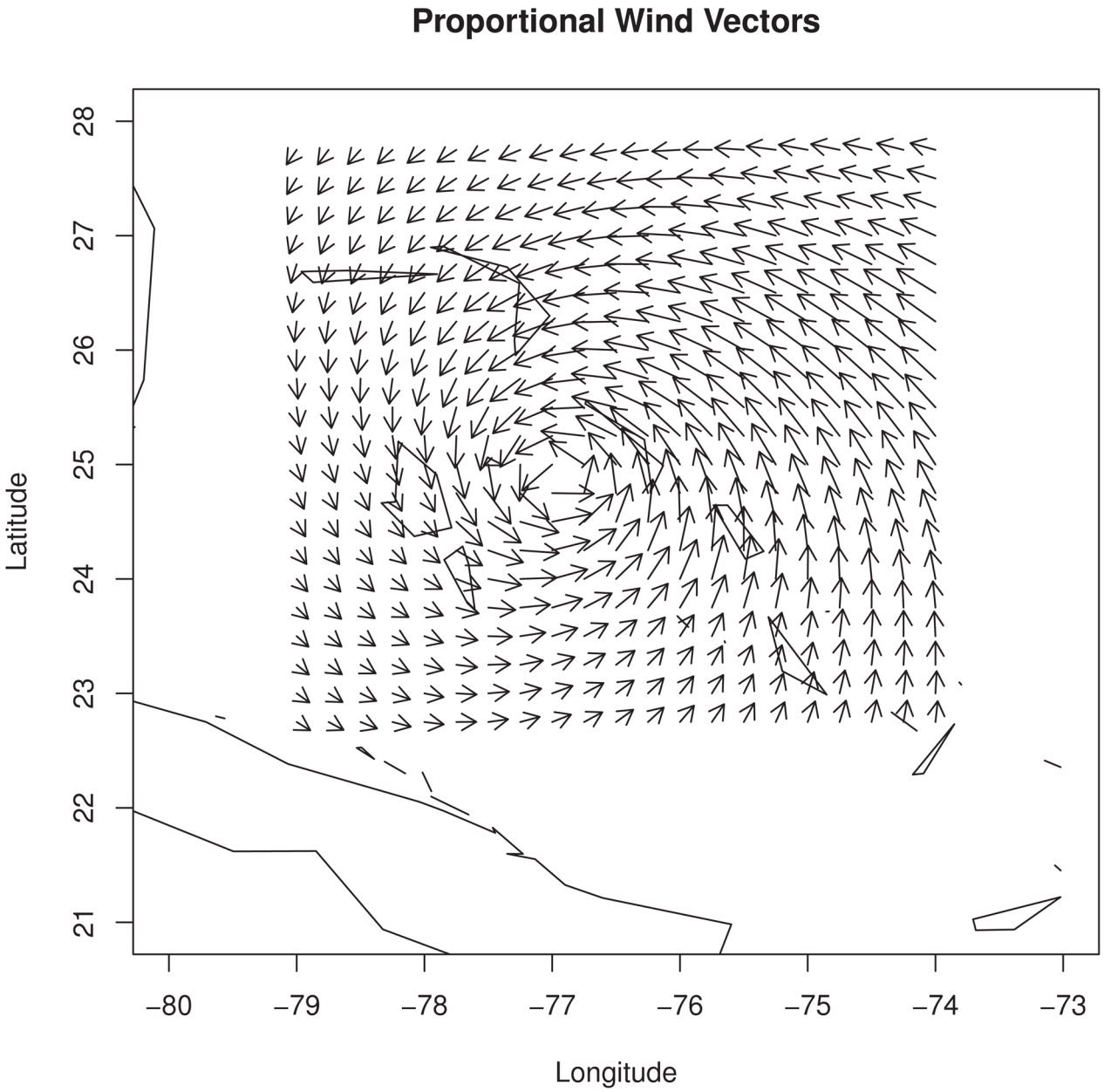


**Figure 1.**
Plot of wind vectors of Hurricane Floyd (09/14/1999 at noon local time).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

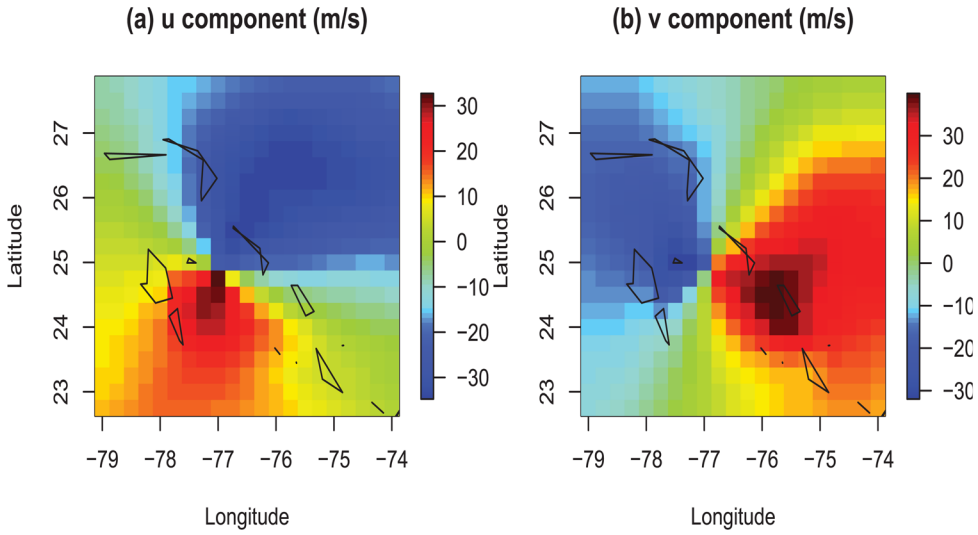


**Figure 2.**
(a) u component. (b) v component.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

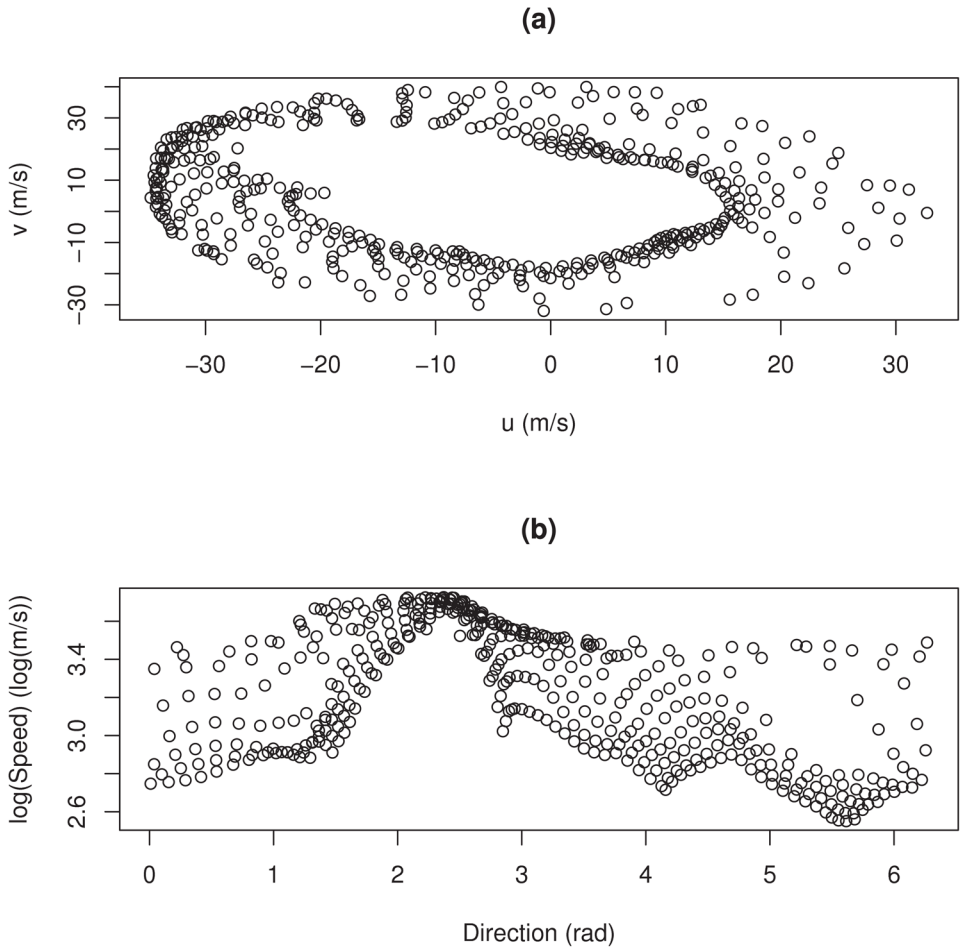


**Figure 3.**
Scatterplots of Hurricane Floyd data: (a) u and v components, (b) log(speed) ($y$) and direction ($\theta$).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

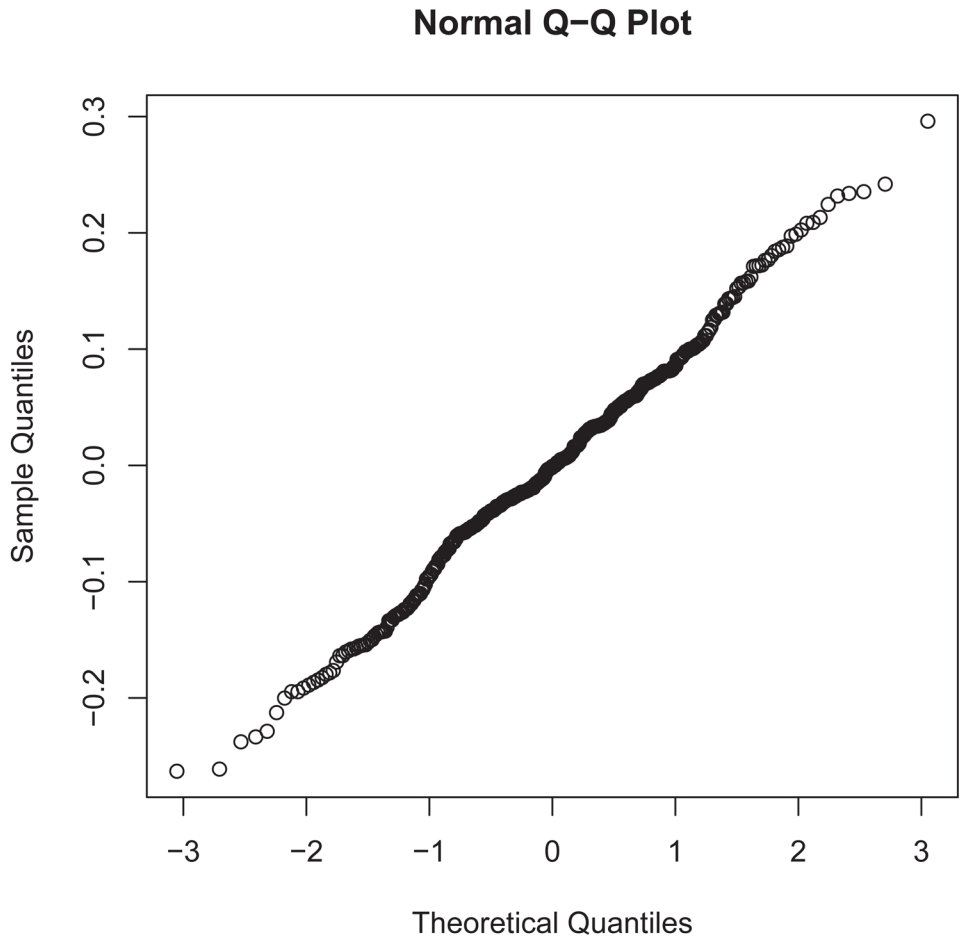


**Figure 4.**
Normal q-q plot of residuals under log transformation of wind speed.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

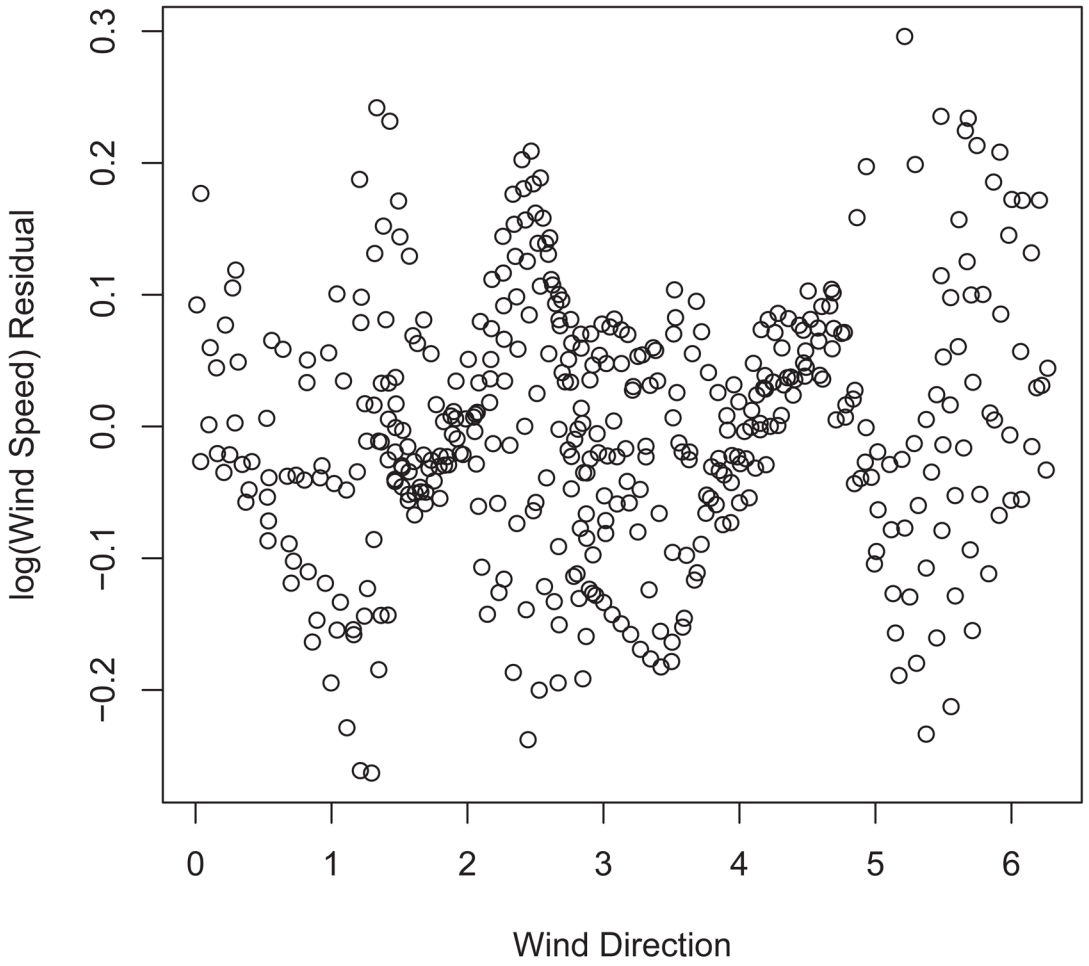


**Figure 5.**
Scatterplot showing wind direction with residuals of model under log transformation of wind speed.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

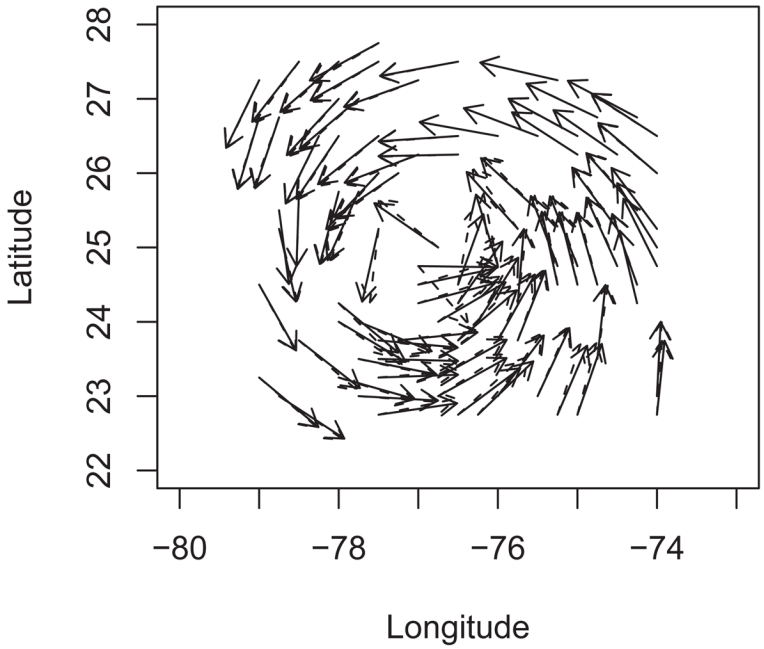


**Figure 6.**
Posterior mean direction, dashed arrow, compared to observed direction, solid arrow.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

**Table 1**

Five-fold cross validation comparison of the U/V model to our CCAR model. Wind speed is compared under the MSE metric while wind direction uses MCDE and MACE. Standard errors are in parentheses.

| Model | MSE($\omega$) | MCDE | MACE |
|---|---|---|---|
| U/V | 29.89 (3.20) | 0.079 (0.003) | 0.983 (0.002) |
| CCAR | 0.96 (0.11) | 0.024 (0.003) | 0.997 (0.001) |

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

**Table 2**

Results of sensitivity analysis of prior selection for variance parameters using angle comparison metrics. Standard errors are in parentheses.

| Prior Selection | MCDE | MACE |
|---|---|---|
| Gamma(0.5,0.0005) | 0.052 (0.027) | 0.763 (0.016) |
| Uniform(0,100) | 0.057 (0.023) | 0.748 (0.017) |
| Uniform(0,300) | 0.046 (0.022) | 0.753 (0.013) |