Dynamic Bayesian temporal modeling and forecasting of short-term wind measurements

Irene García[a], Stella Huo[b], Raquel Prado[c,*], Lelys Bravo[d]

[a]*Departamento de Ciencias Matemáticas e Informática, Universitat de les Illes Balears*
[b]*Division of Pulmonary and Critical Medical Care, Zuckerberg San Francisco General Hospital and Trauma Center, University of California San Francisco*
[c]*Department of Statistics, University of California Santa Cruz*
[d]*Department of Statistics, University of Illinois Urbana-Champaign*

Abstract

We present a new Bayesian modeling approach for joint analysis of wind components and short-term wind prediction. This approach considers a truncated bivariate matrix Bayesian dynamic linear model (TMDLM) that jointly models the u (zonal) and v (meridional) wind components of observed hourly wind speed and direction data. The TMDLM takes into account calm wind observations and provides joint forecasts of hourly wind speed and direction at a given location. The proposed model is compared to alternative empirically-based time series approaches that are often used for short-term wind prediction including the persistence method (naive predictor), as well as univariate and bivariate ARIMA models. Model performance is measured predictively in terms of mean squared errors associated to 1-hour and 24-hour ahead forecasts. We show that our approach generally leads to more accurate short term predictions than these alternative approaches in the context of analysis and forecasting of hourly wind measurements in 3 locations in Northern California for winter and summer months.

Keywords: Bayesian dynamic linear models, matrix-variate dynamic models, joint wind speed and direction forecasts, short-term wind prediction.

*Corresponding author

1. Introduction

Wind speed forecasting has been studied extensively in recent years as it is an essential element in the management of wind power generation [1]. Short-term forecasting of wind speed and other related measurements usually refers to predictions from minutes to days ahead (typically no more than 48 hours), while long-term forecasting deals with predictions of several days, weeks and months ahead [see, e.g., 2]. In this paper we focus on short-term wind forecasting. More specifically, we consider joint modeling and forecasting of hourly wind speed and direction.

Models for wind speed forecasting can be roughly divided into those based on physical models, those based on statistical models and hybrid approaches that combine physical and statistical models [1]. Generally, physical models are preferable when dealing with large scale data and long-term predictions, while statistical models are preferable for short-term forecasting. Statistical approaches are usually based on tools for time series analysis such as ARIMA (autoregressive integrated moving average) models, neural networks, functional regression analysis, state-space models and regime switching models among others [see for example 3, 2, 4, 5]. These approaches can lead to relatively accurate short-term forecasts of wind speed, however, they do not provide forecasts of wind direction. In terms of joint forecast of wind speed and direction, a number of approaches are available. [6] considers methods based on ARMA and VAR (vector autoregressive) models and finds that VAR models can outperform ARMA models on wind lateral and longitudinal components in terms of mean absolute error (MAE) when the correlation between speed and direction is modestly significant. However, models in [6] are not appropriate for dealing with non-stationary wind data, which is the type of wind data analyzed here. [7] proposes non-linear methods based on neural networks for wind speed and direction forecasting. Such methods are tested on wind speed and direction data from public records of the Nevada department of transportation's road weather information system. A time interval of 10 min was used to train and test the methods. Analysis and forecasting of speed and direction were done separately, assuming that these two measurements were independent. This approach is shown to compare favorably against other methods such as echo state networks and methods based on adaptive neuro-fuzzy inference in terms of very short term forecasting. [8] considers a non-parametric kernel density estimation method, and a non-parametric version of the Johnson and Wehrly model [9] to jointly analyze and forecast wind speed and direction. The appeal of non-parametric approaches is that they do not assume any particular distribution form and so, they are usually more flexible for describing the data than parametric models. However, non-parametric methods tend to be much more computationally intensive than parametric methods and often unfeasible in pratical settings. [8] shows that their proposed methods lead to a better performance than alternative parametric models when jointly modeling wind speed and direction data, but no assessment of the quality of short-term forecasts is provided. In addition, non-parametric methods assume bandwidths and other tuning parameters to be fixed over time which may not offer enough flexibility for analyzing data with time-varying features. Finally, we note that none of the methods just mentioned explicitely consider modeling calm winds (i.e., winds with zero speed). This is important as wind data with high temporal resolution (i.e., hourly or more frequent measurements) typically contain a large number of zero observations corresponding to measurements during calm periods of zero wind speed.

In this paper we present a Bayesian model for joint analysis and short-term forecasting of non-stationary wind speed and direction data. We propose, implement and test a truncated bivariate matrix Bayesian dynamic linear model that jointly models the u (zonal) and v (meridional) wind components of observed hourly wind speed and direction data. We note that univariate truncated dynamic linear models have been used to model and forecast rainfall data in [10]. The truncated dynamic linear model presented here is a bivariate generalization of the model in [10]. Our model is able to provide joint forecasts of hourly wind speed and direction. We test our models by analyzing and forecasting median hourly wind speed and direction data from 3 locations in Northern California. These are public data available at the Iowa Environmental Mesonet (IEM) Automated Surface Observing System (ASOS) Network. Model performance is measured predictively in terms of mean squared errors associated to 1-hour and 24-hour ahead forecasts.

The paper is organized as follows. Section 2 presents a description of the data. Section 3 presents the proposed model and discusses algorithms for posterior inference and forecasting. Section 4 shows the data analysis and forecasting with the proposed models as well as comparisons with other approaches. Finally, Section 5 presents final remarks and discusses possible future extensions.

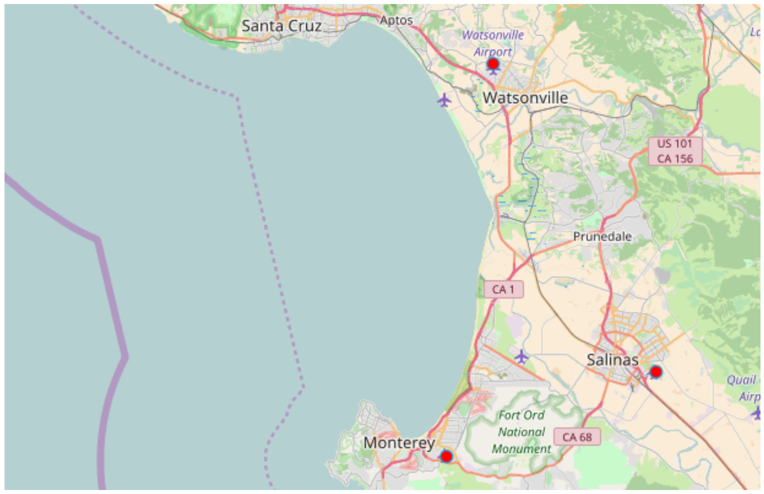


Figure 1: Location of the 3 stations in Northern California.

2. Wind data

Wind data were obtained from the Iowa Environmental Mesonet (IEM) Automated Surface Observing System (ASOS) Network, a publicly available database (`http://mesonet.agron.iastate.edu/ASOS`). ASOS stations are located at airports and take minute-by-minute observations and general basic weather reports for the National Weather Service (NWS), the Federal Aviation Administration (FAA), and the Department of Defense (DOD). These observations are nationally monitored for quality 24 hours per day. For additional detailed information about the ASOS measurements see [11].

For this paper we consider wind direction (in degrees relative to the north) and speed (in knots) from 3 ASOS stations in Northern California near the Monterey Bay Area, specifically, stations located in airports in Watsonville (WVI), Salinas (SNS) and Monterey (MRY) (see Figure 1). Wind direction is reported to the nearest 10 degree increment (e.g., 274 degrees is reported as 270 degrees). The ASOS wind sensors' starting threshold for response to wind direction and speed is 2 knots and so, winds measured at 2 knots or less are reported as calm (i.e., 0 speed magnitude). We consider hourly median wind speed magnitude and corresponding direction for the months of February and August. For illustration purposes we present analyses for these two months for two years –namely, 2010 and 2013– at the three locations listed above. Similar results in terms of the performance of our proposed model were obtained from analyzing data from a 10 year period from 2008 to 2017. We chose the months of February and August for a number of reasons. February is, on average, one of the months with the largest average rainfall and the largest average wind speed magnitude for the three selected locations in Northern California, while August is, on average, one of the months with the lowest average rainfall (essentially none) and the lowest average wind speed magnitude. Also, the wind directions are, on average, very different for these two months. In this paper we evaluate the goodness of fit and forecasting capabilities of our proposed models for these two different months. Figure 2 shows the windrose plots of median hourly wind speed and direction data for the months of January and August from 2008 to 2017 in Monterey, Salinas and Watsonville. Clearly, there are differences across the 3 locations and also seasonal differences. All the locations, specially Salinas, register a larger count of stronger winds (above 17 knots) in the month of February. These winds are from the South-East and some come from the West in Monterey and Salinas, and from the South in Watsonville. In August the winds come mostly from the West, including readings from the South-West and the North-West in Monterey and Salinas, and mostly from the South in Watsonville. Finally, we also considered median hourly air temperature (in ° Farenheit) and sea-level pressure (in mb) as possible covariates in our model.

3. Bayesian dynamic modeling

We propose a Bayesian dynamic model for analysis and forecasting of hourly median wind data for each month, year and location. Our model takes into account wind speed magnitude and direction by jointly modeling the zonal and the meridional components, denoted as u and v components, respectively. The u component is the component towards the East, while the v component is the component towards the North. More specifically, let $\mathbf{y}_t = (y_{t,1}, y_{t,2})'$ for $t = 1 : T$ be a 2-dimensional time series comprising the u (zonal) component and the v (meridional) component of the wind measurement at time t (with t indexing hourly data) for a given location and month, i.e., $y_{t,1} = -s_t \sin(\pi d_t/180)$, and $y_{t,2} = -s_t \cos(\pi d_t/180)$, where s_t is the wind speed in knots and d_t is the meteorological wind direction in degrees clockwise from the north at time t.

We consider the following bivariate truncated dynamic linear model for

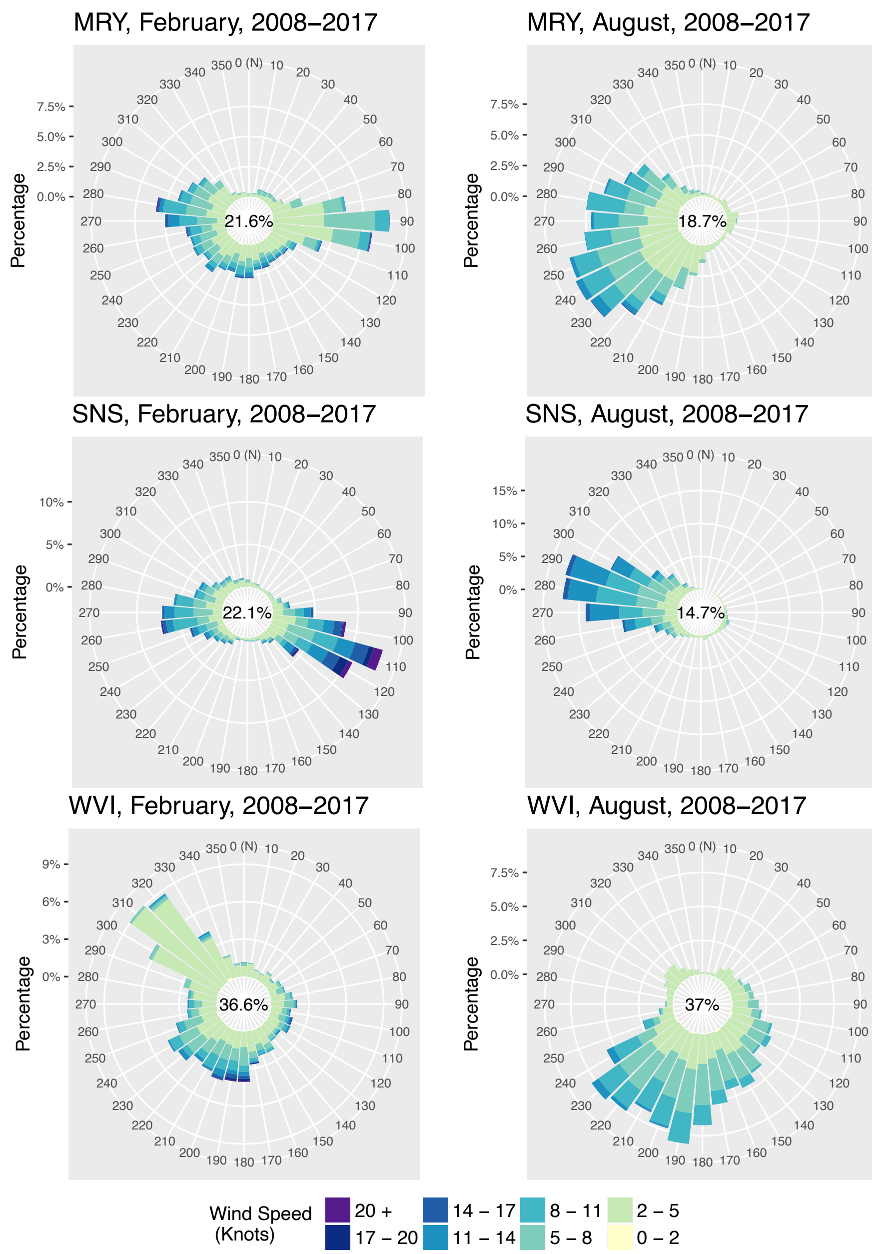


Figure 2: Windrose plots of median hourly wind data for January and August from 2008 to 2017 in Monterey (MRY), Salinas(SNS) and Watsonville (WVI). Numbers in the center correspond to percentages of calm wind (i.e., winds with 0 speed magnitude).

$t = 1 : T$,

$$\mathbf{y}_t \;=\; \begin{cases} (0,0)' & \text{if} \;\; -a \leq h_{t,1} \leq a, \;\; \text{and} \;\; -b \leq h_{t,2} \leq b, \\ \mathbf{h}_t & \text{otherwise,} \end{cases}$$

with a and b fixed values that capture the approximate censoring of the wind measuring devices, and

$$\mathbf{h}_t' \;=\; \mathbf{F}_t'\Theta_t + \boldsymbol{\epsilon}_t', \quad \boldsymbol{\epsilon}_t \sim N_2(\mathbf{0}, v\Sigma), \tag{1}$$

$$\Theta_t \;=\; \mathbf{G}\Theta_{t-1} + \Omega_t, \quad \Omega_t \sim MN_{p\times 2}(\mathbf{0}, \mathbf{W}_t, \Sigma), \tag{2}$$

with initial distributions $(\Theta_0, \Sigma|\mathcal{D}_0) \sim MNW^{-1}_{p\times 2}(\mathbf{m}_0, \mathbf{C}_0, n_{0,\Sigma}, \Sigma_0)$, and $(v|\mathcal{D}_0) \sim IG(n_{0,v}, d_{0,v})$, where MN denotes the matrix normal distribution, W^{-1} denotes the inverse-Wishart distribution, MNW^{-1} denotes the matrix normal inverse-Wishart distribution, and IG denotes the inverse-gamma distribution. Here we have that

- $\mathbf{h}_t = (h_{t,1}, h_{t,2})'$ is a latent process describing the underlying behavior of the 2 hourly wind components over time,

- Θ_t a $p \times 2$ is matrix of state parameters; p denotes the dimension of the parameter space, which depends on the structure of the model and the number of covariates included in the analysis as explained below,

- $\mathbf{F}_t$ is a p-dimensional vector of constants, and $\mathbf{G}$ is a $p \times p$ known state evolution matrix,

- $\boldsymbol{\epsilon}_t$ is a 2-dimensional vector of observational errors, Ω_t is a $p \times 2$ evolution error matrix, assumed to be zero mean matrix-normally distributed, with left $p \times p$ variance matrix $\mathbf{W}_t$ and right 2×2 variance matrix Σ; note that the matrix-normal inverse Wishart prior implies that $(\Sigma|\mathcal{D}_0) \sim W^{-1}(n_{0,\Sigma}, \mathbf{S}_0)$.

Equations (1) and (2) above define a matrix dynamic linear model [see, 12, 13]. Therefore, our bivariate model is a truncated model with an underlying multivariate dynamic linear structure. Truncated univariate dynamic linear models have been used before for analyzing rainfall data in [10]. The model proposed here is in this sense a generalization of [10] to the bivariate case, and we use it for joint modeling and forecasting of short term wind speed magnitude and direction. The value of p and specific structure of $\mathbf{F}_t$ and $\mathbf{G}$

in the context of our models for hourly wind components is detailed below in Section 3.1.

In addition, $\mathbf{W}_t$ is specified sequentially using discount factors as described in [12], i.e., we assume

$$\mathbf{W}_t = \Delta^{-1/2}\mathbf{G}\mathbf{C}_{t-1}\mathbf{G}'\Delta^{-1/2} - \mathbf{G}\mathbf{C}_{t-1}\mathbf{G}', \quad (3)$$

with $\Delta = \text{diag}(\delta_1, \ldots, \delta_p)$ and discount factors $\delta_i \in (0, 1]$ for all $i = 1 : p$. Optimal values of δ_i will be chosen to maximize likelihood-based criteria or to minimize mean squared errors for one-step ahead forecasts.

Posterior inference is achieved via Markov chain Monte Carlo (MCMC) by iteratively sampling from the conditional distributions described below. Given initial values for $n_{0,v}, d_{0,v}, n_{0,\Sigma}, \mathbf{S}_0, \mathbf{m}_0, \mathbf{C}_0, \Theta^0_{1:T}, \Sigma^0$ and setting $h^0_{1,t} = y_{t,1}$ and $h^0_{2,t} = y_{t,2}$ for all t, at iteration i we proceed as follows:

- Draw $v^{(i)}$ from $IG(n_{T,v}/2, d_{T,v}/2)$, with $n_{T,v} = n_{0,v} + 2T$ and

$$d_{T,v} = d_{0,v} + \sum_{t=1}^{T} [(\mathbf{h}_t^{(i-1)})' - \mathbf{F}'\Theta_t^{(i-1)}](\Sigma^{(i-1)})^{-1}[(\mathbf{h}_t^{(i-1)})' - \mathbf{F}'\Theta_t^{(i-1)}]'.$$

- Draw $(\Theta^{(i)}_{1:T}, \Sigma^{(i)})$ using the algorithm of [14], which combines the forward-filter-backward-sampling (FFBS) of [15] and [16] with the algorithm of [17]. This is done as follows:

  – (Forward Filtering) for $t = 1 : T$, compute $\mathbf{a}_t, \mathbf{R}_t, \mathbf{f}_t, Q_t, \mathbf{m}_t, \mathbf{C}_t$, $n_{t,\Sigma}$, and $\mathbf{S}_t$, as

$$\begin{aligned} \mathbf{a}_t &= \mathbf{G}\mathbf{m}_{t-1}, & \mathbf{R}_t &= \mathbf{G}\mathbf{C}_{t-1}\mathbf{G}' + \mathbf{W}_t, \\ \mathbf{f}'_t &= \mathbf{F}'\mathbf{a}_t, & Q_t &= \mathbf{F}'\mathbf{R}_t\mathbf{F} + v, \\ \mathbf{m}_t &= \mathbf{a}_t + \mathbf{A}_t\mathbf{e}'_t, & \mathbf{C}_t &= \mathbf{R}_t - \mathbf{A}_t Q_t \mathbf{A}'_t, \end{aligned}$$

  with $\mathbf{e}_t = \mathbf{h}_t^{(i-1)} - \mathbf{f}_t$, $\mathbf{A}_t = \mathbf{R}_t\mathbf{F}_t/Q_t$, $n_{t,\Sigma} = n_{t-1,\Sigma} + 1$, and

$$\mathbf{S}_t = \frac{1}{n_{t,\Sigma}} \left(n_{t-1,\Sigma}\mathbf{S}_{t-1} + \mathbf{e}_t\mathbf{e}'_t/Q_t \right).$$

  Again, note that $\mathbf{W}_t$ is specified via (3).

  – Sample $\Sigma^{(i)}$ from the distribution $W^{-1}(n_{T,\Sigma}, \mathbf{S}_T)$.

– (Backward Sampling) Sample $\Theta_T^{(i)}$ from $MN_{p\times 2}(\mathbf{m}_T, \mathbf{C}_T, \Sigma^{(i)})$, and then, for $t = (T-1) : 0$, sample $\Theta_t^{(i)}$ from $MN_{p\times 2}(\mathbf{m}_t^*, \mathbf{C}_t^*, \Sigma^{(i)})$, with

$$\begin{aligned} \mathbf{m}_t^* &= \{\mathbf{I}_p - \mathbf{C}_t\mathbf{G}'\mathbf{R}_{t+1}^{-1}\mathbf{G}\}\mathbf{m}_t + \mathbf{C}_t\mathbf{G}'\mathbf{R}_{t+1}^{-1}\Theta_{t+1}^{(i)}, \\ \mathbf{C}_t^* &= \{\mathbf{I}_p - \mathbf{C}_t\mathbf{G}'\mathbf{R}_{t+1}^{-1}\mathbf{G}\}\mathbf{C}_t, \end{aligned}$$

where $\mathbf{I}_p$ denotes the identity matrix of dimension p.

- For $t = 1 : T$ sample $\mathbf{h}_t^{(i)}$ as follows. For each t, if $\mathbf{y}_t \neq \mathbf{0}$, set $\mathbf{h}_t^{(i)} = \mathbf{y}_t$, otherwise sample $\mathbf{h}_t^{(i)}$ from a bivariate truncated normal $TN_{(-a,a)\times(-b,b)}(\mathbf{F}_t'\Theta_t^{(i)}, v^{(i)}\Sigma^{(i)})$.

3.1. Specific model structure

The model proposed above is very general in the sense that $\mathbf{F}_t$, $\mathbf{G}$ can be specified by the modeler to include trend, seasonal components, and any additional covariates. We now describe the structure used in our analyses of the wind component data for the 3 locations and each of the months. As mentioned above, we consider two covariates, namely, $x_{1,t}$:= the air temperature in ° Farenheit and $x_{2,t}$:= the mean sea level pressure (in mb). We also consider seasonal components by using a Fourier DLM representation [12] with fundamental period 24 for the hourly data. Then, the complete Fourier model with a fundamental period of 24, all the harmonics, and the two covariates listed above, is a model with $p = 25$, a 25×2 matrix of state parameters, and $\mathbf{F}_t$ and $\mathbf{G}$ given by:

$$\begin{aligned} \mathbf{F}_t &= (x_{1,t}, x_{2,t}, \mathbf{E}_2', \ldots, \mathbf{E}_2', 1)', \\ \mathbf{G} &= \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \mathbf{0} & \mathbf{0} & J_2(1,\omega) & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & J_2(1,2\omega) & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & J_2(1,(h-1)\omega) & \mathbf{0} \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 \end{pmatrix}, \end{aligned}$$

with $\omega = 2\pi/24$, $h = 12$, $\mathbf{E}_2 = (1, 0)'$, and

$$J_2(1, r\omega) = \begin{pmatrix} \cos(\omega r) & \sin(\omega r) \\ -\sin(\omega r) & \cos(\omega r) \end{pmatrix},$$

for $r = 1, \ldots, 11$. Therefore, $\mathbf{F}_t$ in this case is a 25-dimensional vector and $\mathbf{G}$ is a 25×25 matrix. In general, only a few harmonics of the fundamental period are needed, leading to more parsimonious representations with smaller p.In other words, we can consider smaller models that include only a subset of the entire set of harmonics of the fundamental period. For instance, a model with 5 harmonic components and 2 covariates has $p = 12$. We assess the importance of the harmonics by computing highest posterior density regions (HPDs) and corresponding probabilities of retention at time T for individual harmonics as proposed in [12].

Regarding the specification of $\mathbf{W}_t$, we use discount factors as mentioned above. We consider 3 different discount factors: one discount factor for each of the 2 covariates, namely, δ_1 and δ_2, and an additional discount factor for the seasonal components, denoted as δ_S. Then, in the full seasonal model for hourly data, Δ is a 25×25 matrix with $\Delta = diag(\delta_1, \delta_2, \delta_S, \ldots, \delta_S)$.

4. Data analysis and results

We fit the truncated bivariate model described above to the wind components for each of the 3 locations and each of the months considered in this analysis, namely, February and August of 2010 and February and August of 2013. We began by selecting the number of significant harmonics and the optimal discount factors in each case. The number of harmonics was determined as explained in [12] by computing the probabilities of retention at the last observed point T for each individual component. Based on these results we determined that, for most locations, months, and years, at most the first 5 harmonics were significant. We also looked at the predictions from models that used a number of harmonics larger than 5, however, we found no substantial improvements in terms of the 24-hours ahead predictions and goodness of fit measurements. Therefore, and specially in order to provide comparisons across different years and locations, we used models that used only the first 5 harmonics in all cases. Note that this results in a dimension reduction of the truncated bivariate DLM from a 25-dimensional state parameter vector in the case of the complete Fourier seasonal model with the 2 additional covariates (temperature and pressure), to a reduced model with a 12-dimensional state parameter vector ($p = 12$) that also includes the 2 covariates.

Regarding the discount factors, we considered a grid of values for δ_1, δ_2 and δ_S in $(0.9, 1] \times (0.9, 1] \times (0.9, 1]$, and chose the optimal values that mimimized the mean squared errors (MSEs) of the one-hour ahead predictions in each case.

We also considered 4 versions of our proposed model: (a) the original version that includes both, temperature and pressure as covariates; (b) a version that includes only temperature; (c) a version that includes only pressure and (d) a version with no covariates. Table 1 compares the 3 versions of the truncated bivariate matrix DLM model in terms of the MSE for the 1-hour ahead and 24-hours ahead forecasts for these models. For both months in 2013 we see that, among our truncated bivariate matrix DLM models, the model that includes only pressure as covariate is either the one with the smallest 24-hour ahead MSE values for all the locations, or it leads to MSE values that are similar to those obtained with other models. The only exception being February 2013 in Salinas for which the model with all the covariates produces a much smaller MSE for the 24-hours ahead prediction. We also see that Watsonville has smaller MSEs than the rest of the locations across all the models, indicating that our model does best at predicting wind components in this location. Finally, we note that the model with no covariates does substantially worse in terms of the MSEs for most locations and months. Similar results were obtained from the analysis with the months of February and August of 2010, however, due to space limitations we are not including these results here.

In order to show the performance of our models in terms of goodness of fit and prediction, we computed the estimated posterior means as well as the 24-hours ahead forecasts for the wind components, along with corresponding 95% posterior intervals, for the months of February and August of 2010 and 2013 in Monterey, Salinas and Watsonville. Figure 3 shows this posterior fit and short-term forecasts for the Salinas location. We see that our proposed bivariate model adequately captures the 24-hour observed seasonality in the wind data and leads to reasonable estimates and forecasts. Similar results were obtained for the other two locations and years.

Figure 4 shows windrose plots of posterior estimates of wind speed magnitude and direction obtained from transforming the estimated values for wind components obtained from our bivariate truncated dynamic linear model for February and August 2013 in Salinas. Overall we can see that the estimates from the model adequately capture the behavior of the observed wind speed magnitude and direction in these two months at this location. Similar results in terms of the goodness of fit were obtained for the other two locations.

Figures 5 and 6 provide a more detailed assessment of the quality of the

		MSE (1-hour ahead)			MSE (24-hours ahead)		
		MRY	SNS	WVI	MRY	SNS	WVI
	(a)	**2.56**	**2.83**	**1.90**	6.51	**14.14**	5.03
FEB	(b)	2.75	2.99	2.00	6.02	18.65	**4.08**
2013	(c)	2.80	3.02	1.98	**5.87**	18.05	4.74
	(d)	6.45	6.95	2.87	6.17	18.70	6.51
		MSE (1-hour ahead)			**MSE (24-hours ahead)**		
		MRY	SNS	WVI	MRY	SNS	WVI
	(a)	1.85	2.20	1.52	8.31	10.11	3.13
AUG	(b)	**1.83**	**2.16**	**1.51**	8.42	**9.27**	**2.88**
2013	(c)	1.85	2.18	1.53	**8.23**	9.39	3.11
	(d)	8.50	17.09	4.91	16.70	72.16	9.21

Table 1: MSE values for 1-hour ahead and 24-hours ahead forecasts of the u and v components in February and August 2013 from truncated bivariate dynamic models with (a) both, temperature and pressure as covariates; (b) only temperature; (c) only pressure and (d) no covariates.

short-term forecasts produced by our models. Figure 5 shows the traces of the 24 hour ahead forecasts and corresponding 95% uncertainty bands obtained from our bivariate matrix DLMs for the months of February and August of 2013 for all 3 locations. Figure 6 shows windrose plots of the actual observations and the predictions obtained from our models for all the locations on February 28 2013 and on August 31 2013. We see that in general, the predicted values from our model adequately capture the magnitude of the speed and the direction of the winds for the two periods of 24 hours considered in all the locations.

Finally, it is also possible to obtain posterior inference on the variance-covariance matrix of the error term of the bivariate latent structure of the non-zero wind components Σ in equation (1). The posterior samples of Σ obtained from the MCMC allow us to make inference on the correlation between the u and v components for different months and locations. Table 2 shows the posterior mean and 95% posterior interval for the correlation between the two wind components for each of the months (February and August 2013) at each of the three locations. From this table we see that there is a significant negative correlation between the two wind components for the Salinas location in February 2013, and significant positive correlation between the two wind components in August 2013 for the Monterey and

		Posterior Mean	95% Posterior Interval
FEB 2013	MRY	0.076	$(-0.032, 0.133)$
	SNS	$\mathbf{-0.162}$	$(\mathbf{-0.209}, \mathbf{-0.038})$
	WVI	0.004	$(-0.097, 0.069)$
AUGUST 2013	MRY	**0.192**	$(\mathbf{0.070}, \mathbf{0.241})$
	SNS	0.085	$(-0.009, 0.130)$
	WVI	**0.108**	$(\mathbf{0.000}, \mathbf{0.158})$

Table 2: Posterior estimates of the correlation between the u and v components obtained from the bivariate TMDLM for February and August 2013 in Monterey (MRY), Salinas (SNS) and Watsonville (WVI).

Watsonville locations. The proposed TMDLM model not only provides a way to estimate the correlation between the two wind components, but also takes this correlation into account to produce more accurate short-term forecasts.

4.1. Comparison to other modeling approaches

In this section we compare the performance of our proposed truncated bivariate matrix normal DLM (bivariate TMDLM) to alternative statistical approaches. We compare the following models: (i) our proposed bivariate TMDLM; (ii) a univariate version of the truncated dynamic linear model (TDLM); (iii) bivariate ARIMA models (iv) univariate ARIMA models and (v) the so called persistence method or naive predictor which consists on using the actual observed value at time t as the predicted value at times $t+h$ for $h = 1, \ldots, 24$ [7]. For the bivariate TMDLM, the univariate TDLM, and also for the bivariate and univariate ARIMA we considered models that include 5 harmonics of the fundamental period and either no covariates, only temperature, only pressure and both, pressure and temperature included as covariates. The proposed bivariate ARIMA had order $p = 3$ for the autoregressive component and order $q = 3$ for the moving average component, while the univariate ARIMA had order $p = 2$ and $q = 3$. These model orders were the optimal model orders obtained using AIC. For the TMDLM and the TDLM optimal discount factors were chosen over a grid of values in $(0.9, 1] \times (0.9, 1] \times (0.9, 1]$.

Table 3 shows the mean squared error (MSE) values obtained from the different approaches taking into account both wind components. In the case of the TMDLM, TDLM, and the ARIMA models we are reporting the results only for the type of model that produced the smallest 24-hour ahead MSEs for

	MRY, FEB 2013	MRY, AUG 2013
Bivariate TMDLM	**5.87**	**8.23**
Univariate TDLM	6.08	8.24
Bivariate ARIMA	8.74	9.72
Univariate ARIMA	6.65	9.66
Persistence	60.81	33.90

Table 3: MSE values of the 24-hour ahead prediction errors for the months of February and August 2013 at the Monterey station obtained from the different models.

the TMDLM among the 4 types of models we considered (again, no covariate, only temperature, only pressure, or both temperature and pressure). Note that due to space restrictions we only report the results of the comparison for the Monterey location, however, similar results were obtained for the other two locations. Therefore, based on the results reported in Table 1 we show the results obtained from TMDLM, TDLM, and ARIMA models with 5 harmonics and only pressure as a covariate for February 2013 and August 2013 in Monterey. Overall we see that the bivariate TMDLM, the univariate TDLM and the ARIMA models do a much better job than the naive/persistance predictor method. We also see that the univariate and bivariate truncated dynamic linear models lead to smaller MSEs that the bivariate and univariate ARIMA models, and that bivariate models generally dominate the univariate models. Our proposed bivariate TMDLM leads to the smallest MSE values in terms of short-term (24-hour) forecasts of the wind components for the Monterey location for the two months considered in 2013.

Similarly, Figure 7 provides a comparative assessment of the predictive performance of the proposed bivariate TMDLM model and the bivariate ARIMA model for February 2013 in the MRY location. The plots display the mean squared error (MSE) and the mean absolute error (MAE) obtained for the 6H, 12H, 18H and 24H ahead predictions from these two models. Again, both models use 5 harmonics and pressure as a predictor. The bivariate ARIMA has AR model order 3 and MA model order 3. No ARMA components are used in the TMDLM. The TMDLM has the lowest MSE and also the lowest MAE values for all the predictions, leading to an improved performance with respect to the bivariate ARIMA model. Similar results are obtained for other months and locations.

5. Conclusion

A Bayesian bivariate truncated matrix dynamic linear model (TMDLM) is proposed for joint analysis and forecasting of wind speed magnitude and direction data that also takes into account calm wind observations. Hourly wind data from 3 locations near the Monterey Bay Area in California were analyzed with the proposed model. The results show that the proposed bivariate TMDLM provides good 24-hour ahead forecasts of wind speed and direction for these locations in months with very different wind and environmental patterns. Furthermore, the TMDLM compares very favorably with alternative statistical models that are commonly used in practice for short-term wind prediction, generally producing more accurate short term forecasts. In addition, the proposed truncated bivariate dynamic linear models also allow us to make inferences on quantities that univariate models are not able to estimate and consider for obtaining more accurate prediction, such as the correlation structure between wind components.

Acknowledgments

Irene García's work on this project was supported by the Schlumberger Foundation through its Faculty For the Future Program. Raquel Prado was partially funded by the National Science Foundation grant SES-1853210.

References

[1] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, Z. Yan, A review on the forecasting of wind speed and generated power, Renewable and Sustainable Energy Reviews 13 (2009) 915–920.

[2] Y. Jiang, Z. Song, A. Kusiak, Very short-term wind speed forecasting with Bayesian structural break model, Renewable Energy 50 (2013) 637–647.

[3] O. Shukur, M. Lee, Daily wind speed forecasting through hybrid KF-ANN model based ARIMA, Renewable Energy 76 (2015) 637–647.

[4] B. Doucoure, K. Agbossou, A. Cardenas, Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data, Renewable Energy 92 (2016) 637–647.

[5] K. Chen, J. Yu, Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach, Applied Energy 114 (2014) 690–705.

[6] E. Erdem, J. Shi, ARMA based approaches for forecasting the tuple of wind speed and direction, Applied Energy 88 (2010) 1405–1414.

[7] M. Chitsazan, M. Fadali, A. Trzynadlowski, Wind speed and wind direction forecasting using echo state network with nonlinear functions, Renewable Energy 131 (2019) 879–889.

[8] Q. Han, Z. Hao, T. Hu, F. Chu, Non-parametric models for joint probabilistic distributions of wind speed and direction data, Renewable Energy 126 (2018) 1032–1042.

[9] R. Johnson, T. Wehrly, Some angular-linear distributions and related regression models, Journal of the American Statistical Association 73 (1978) 602–606.

[10] B. Sanso, L. Guenni, A non-stationary multisite model for rainfall, Journal of the American Statistical Association 95 (2000) 1089–1100.

[11] NOAA, DoD, FAA, USNavy, Automated Surface Observing System (ASOS) User's Guide (1998).

[12] M. West, P. J. Harrison, Bayesian Forecasting and Dynamic Models, 2nd Edition, Springer-Verlag, New York, 1997.

[13] R. Prado, M. West, Time Series: Modeling, Computation and Inference, CRC Press, Boca Raton, FL, 2010.

[14] M. Salvador, J. Gallizo, P. Gargallo, A dynamic principal components analysis based on multivariate normal dynamic linear models, Journal of Forecasting 22 (2003) 457–478.

[15] C. K. Carter, R. Kohn, Gibbs sampling for state space models, Biometrika 81 (1994) 541–553.

[16] S. Frühwirth-Schnatter, Data augmentation and dynamic linear models, Journal of Time Series Analysis 15 (1994) 183–202.

[17] S. Chib, E. Greenberg, Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models, Journal of Econometrics 68 (1995) 339–360.

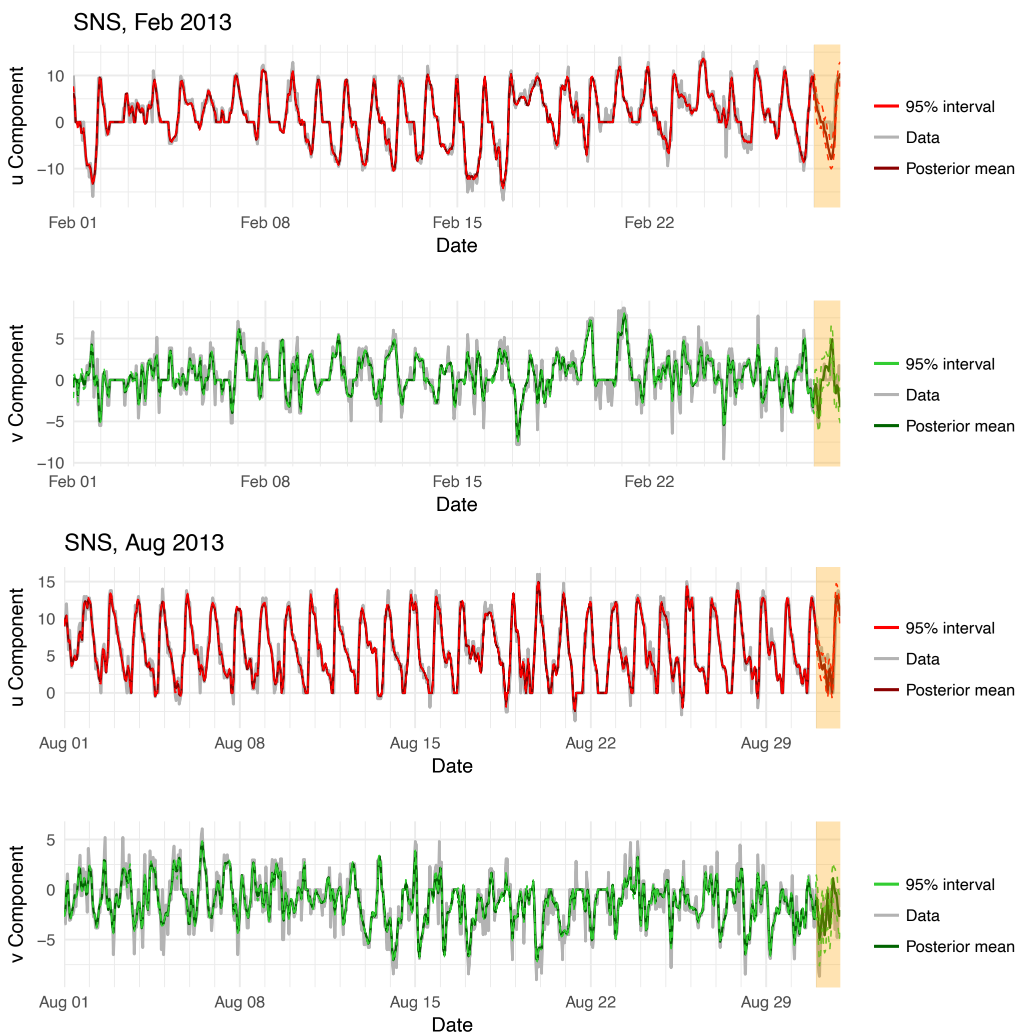


Figure 3: Posterior mean levels, 24 hours ahead forecasts, and corresponding 95% posterior intervals, for the u and v components in February 2013 and August 2013 in Salinas (SNS) obtained from the bivariate matrix DLM with 5 harmonics and temperature and pressure as covariates.

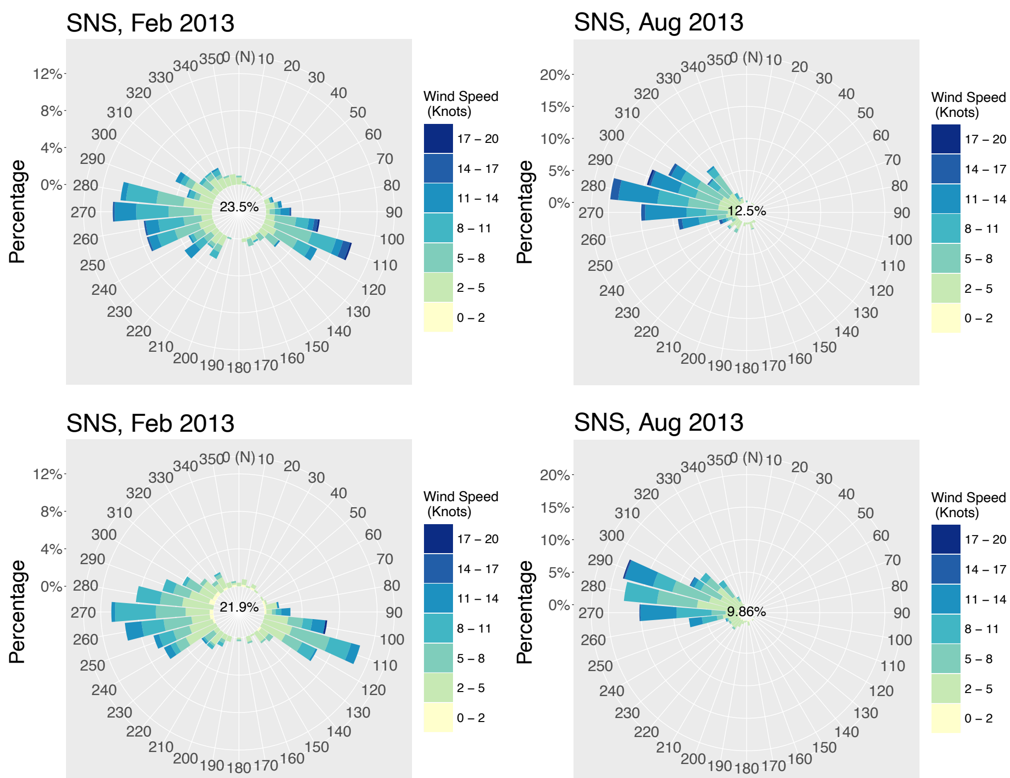


Figure 4: Top: Windrose plots of observed wind speed magnitude and direction in February 2013 (left) and August 2013 (right) in Watsonville. Bottom: Windrose plots of the posterior estimates of wind speed magnitude and direction from the bivariate matrix DLM on the same months and location. The number at the center shows the percentage of calm wind observations.

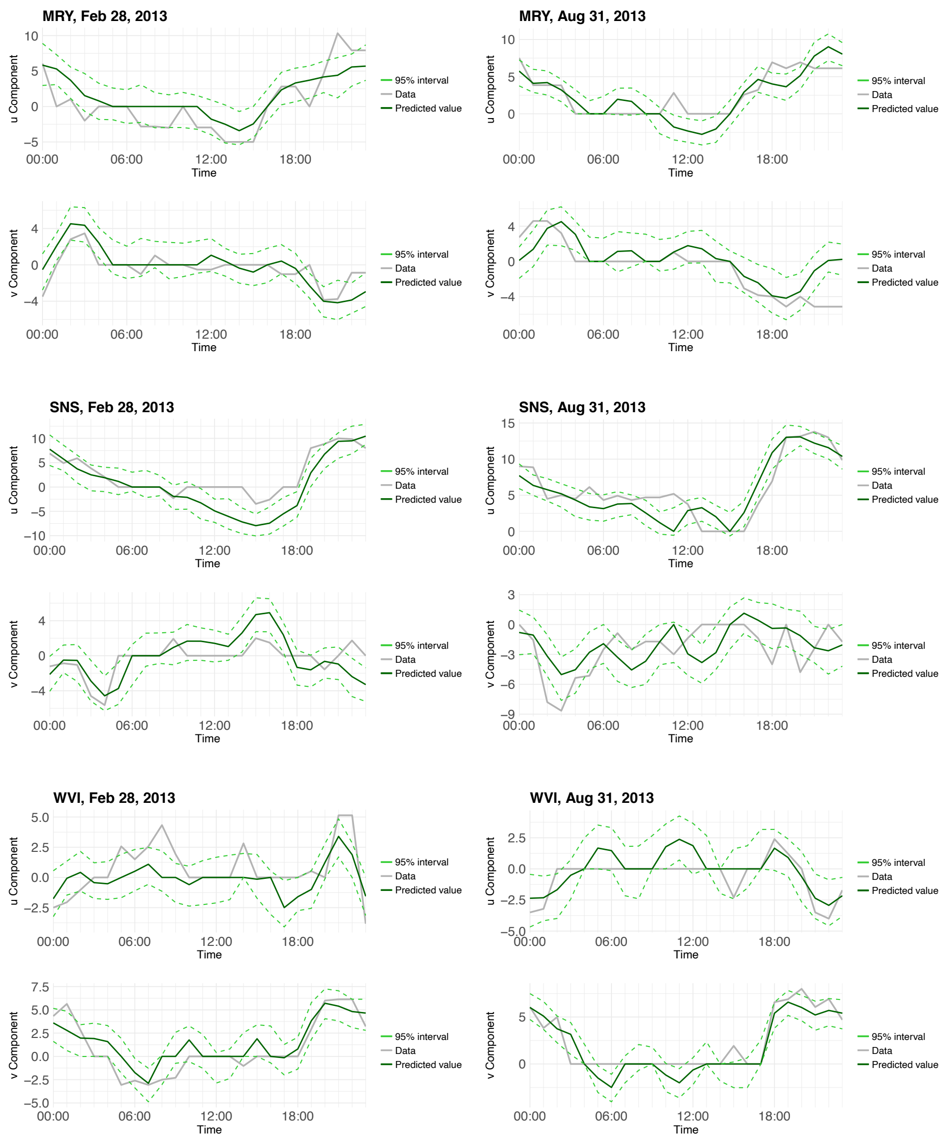


Figure 5: Plots of 24 hours ahead forecast for the u and v components in the last day of February 2013 and August 2013 at Monterey (MRY), Salinas (SNS), and Watsonville (WVI).

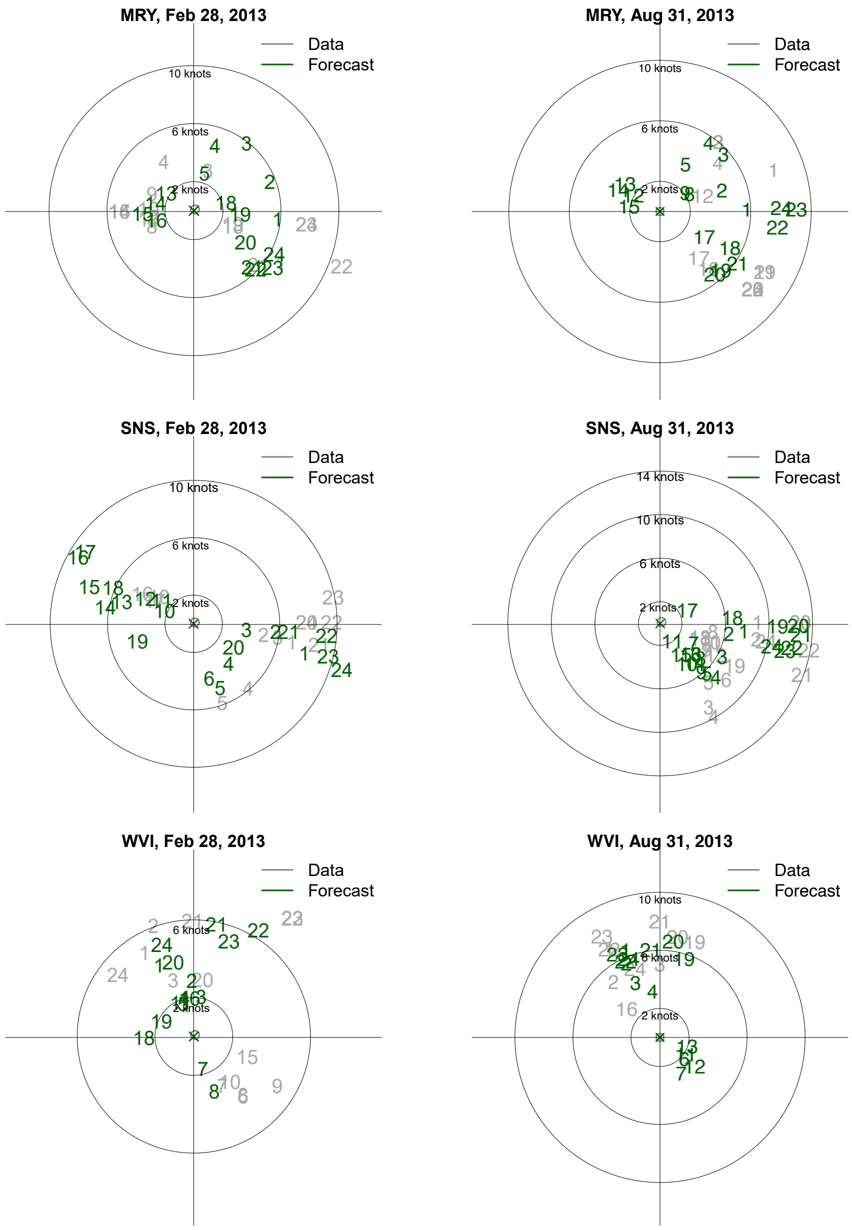
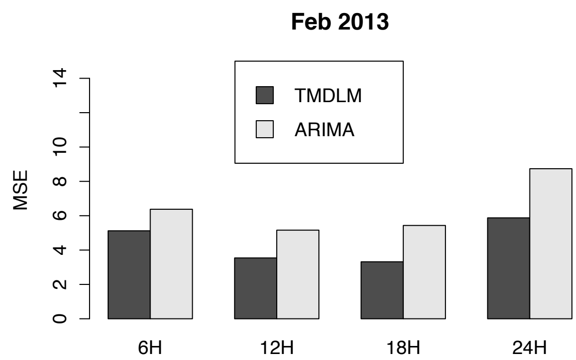
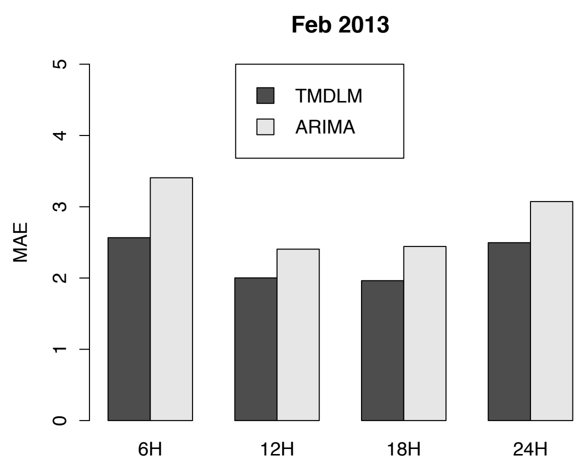


Figure 6: 24-hour ahead forecasts of the speed magnitude and direction (green) along with the actual observations (gray) on February 28 2013 and August 31 2013 in Monterey (MRY), Salinas (SNS), or Watsonville (WVI).



(a)



(b)

Figure 7: MSE (plot (a)) and MAE (plot (b)) from the bivariate TMDLM and ARIMA models for the month of February 2013 in MRY.