Joint modeling of wind speed and wind direction through a conditional approach
Supplementary Material

Eva Murphy*, Whitney Huang†, Julie Bessac‡, Jiali Wang§, Rao Kotamarthi¶

May 29, 2023

SM 1: Estimating the directional wind speed distribution using BWHR with equal bins, BWHR with unequal bins and MLE

We perform a sensitivity analysis for estimating the directional wind speed distribution using the BWHR method with equal bins, unequal (or equal frequency) bins, and MLE. In the case of equal bins, we use $N = 36$ and $K_\alpha = K_\beta = 8$. In the case of unequal bins, we bin the wind direction data so that each bin has the same amount of points. The MLE is done without any binning, where a Weibull distribution is fitted to the entire data using eight pairs of Fourier series to construct the dependence between the shape (and scale) parameter on wind direction. The estimated 50%, 75%, and 95% conditional quantiles for each method are plotted in Fig. 1, suggesting that the BWHR with unequal bins and MLE methods are numerically unstable when the wind direction data is sparse with respect to some directions.

We conclude this section by suggesting to estimate the directional wind speed distribution using the BWHR method with equal bins, i.e. $N = 36$.

*Clemson University. E-mail: nagy@clemson.edu
†Clemson University. E-mail: wkhuang@clemson.edu
‡Argonne National Laboratory. E-mail: jbessac@anl.gov
§Argonne National Laboratory. E-mail: jialiwang@anl.gov
¶Argonne National Laboratory. E-mail: vrkotamarthi@anl.gov

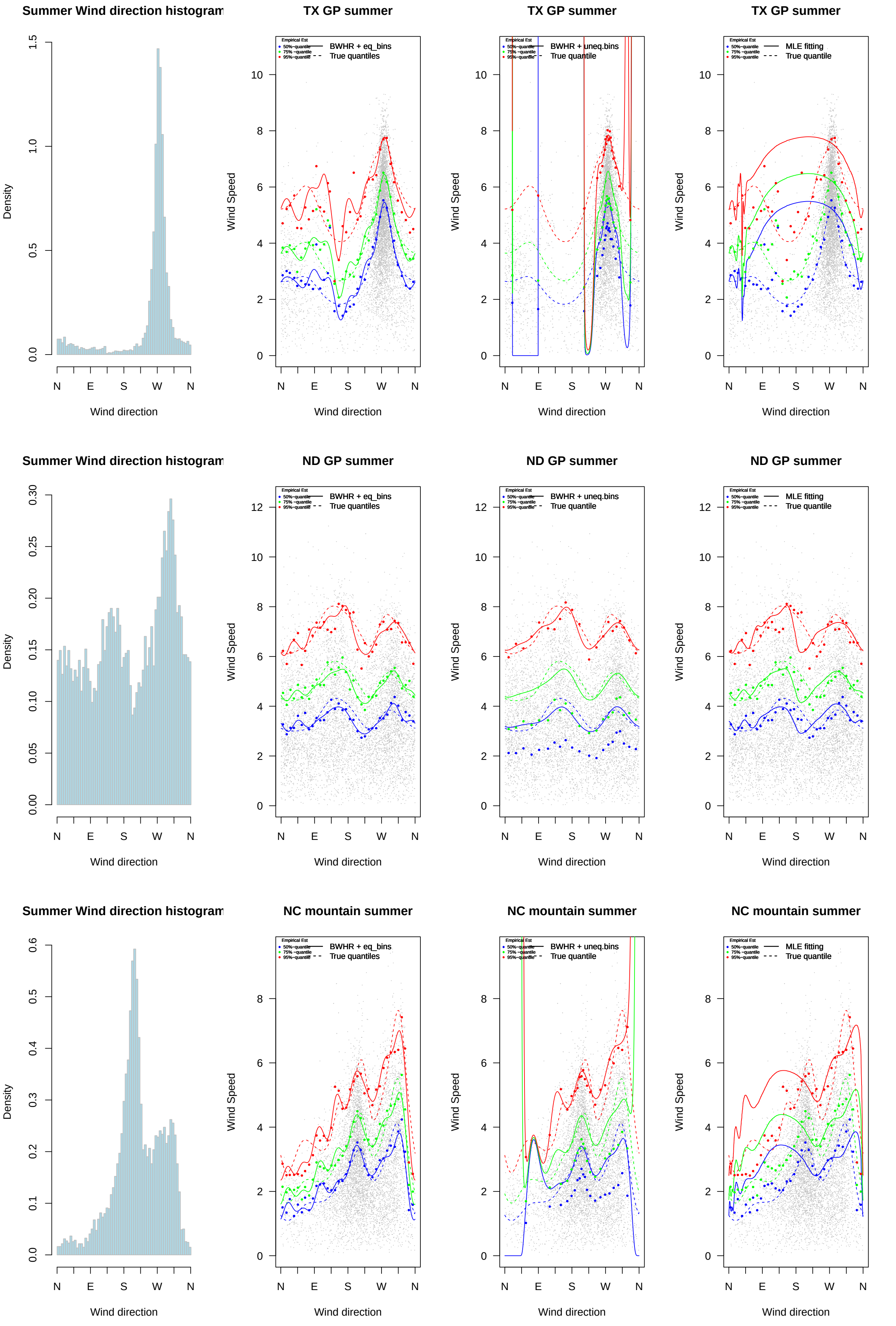

Figure 1: Sensitivity analysis for estimating the quantile curves using the BWHR method with equal bins, unequal bins, and MLE at three different locations. The histograms represent the wind direction distribution at the given location. Each row represents a location.

SM 2: Analysis on the number of bins

We conduct a sensitivity analysis to understand how to bin the wind direction data. We look at different sample sizes ranging from 1000 to 70000 (r, ϕ) points, and use equal bins and unequal bins. In the case of equal bins, we divide the wind direction data so that each bin has a width of 20 degrees (i.e. $N = 18$) and 10 degrees (i.e. $N = 36$), respectively. The unequal bins are constructed such that each bin has the same amount of points and their total number matches the number of equal bins, i.e. $N = 18$ and $N = 36$, respectively. The results of the analysis are plotted in Fig. 2 and they help us understand the followings:

1. When binning the wind direction with unequal bins, some of the bins have a wider width causing the `BWHR` (also referred to as BinWeiHar) to break down (Fig. 1); this is likely because fitting the wind speed data to a Weibull distribution in these large bins may be inappropriate.

2. When binning sparse wind data in some wind directions using equal bins, the number of bins matters. Specifically, having $N \geq 36$ can cause some bins to have no to small data points, in which case the MLE optimizer might fail to converge. On the other hand, having $N < 18$ can lead to wider bins causing issues similar to that of unequal bins.

3. In the case of large sample sizes, i.e $n \geq 70000$, using $N = 36$ bins is already enough as shown in Fig. 3.

We conclude this analysis by suggesting equal bins to avoid having the `BWHR` method break down. Furthermore, the bins should be constructed to include, on average, at least 200 points (i.e., $N = 18$ or $N = 36$).

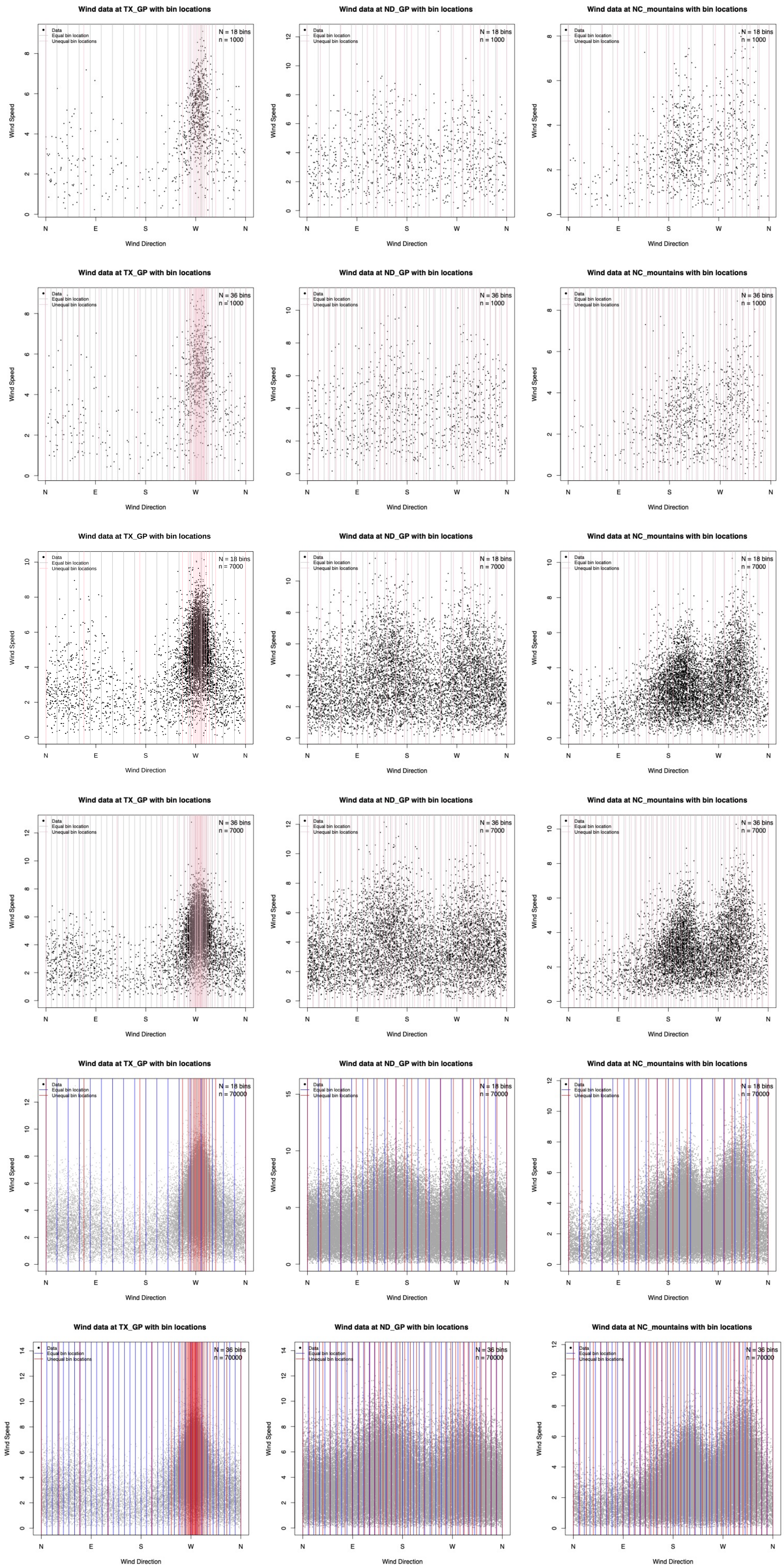

Figure 2: The wind direction domain is binned using equal bins (gray/blue lines) and unequal bins (pink/red lines) using N = 18 bins (rows 1, 3, 5) and N=36 bins (rows 2, 4, 6). Each column represents a different location.Every two rows represent a different sample size n.

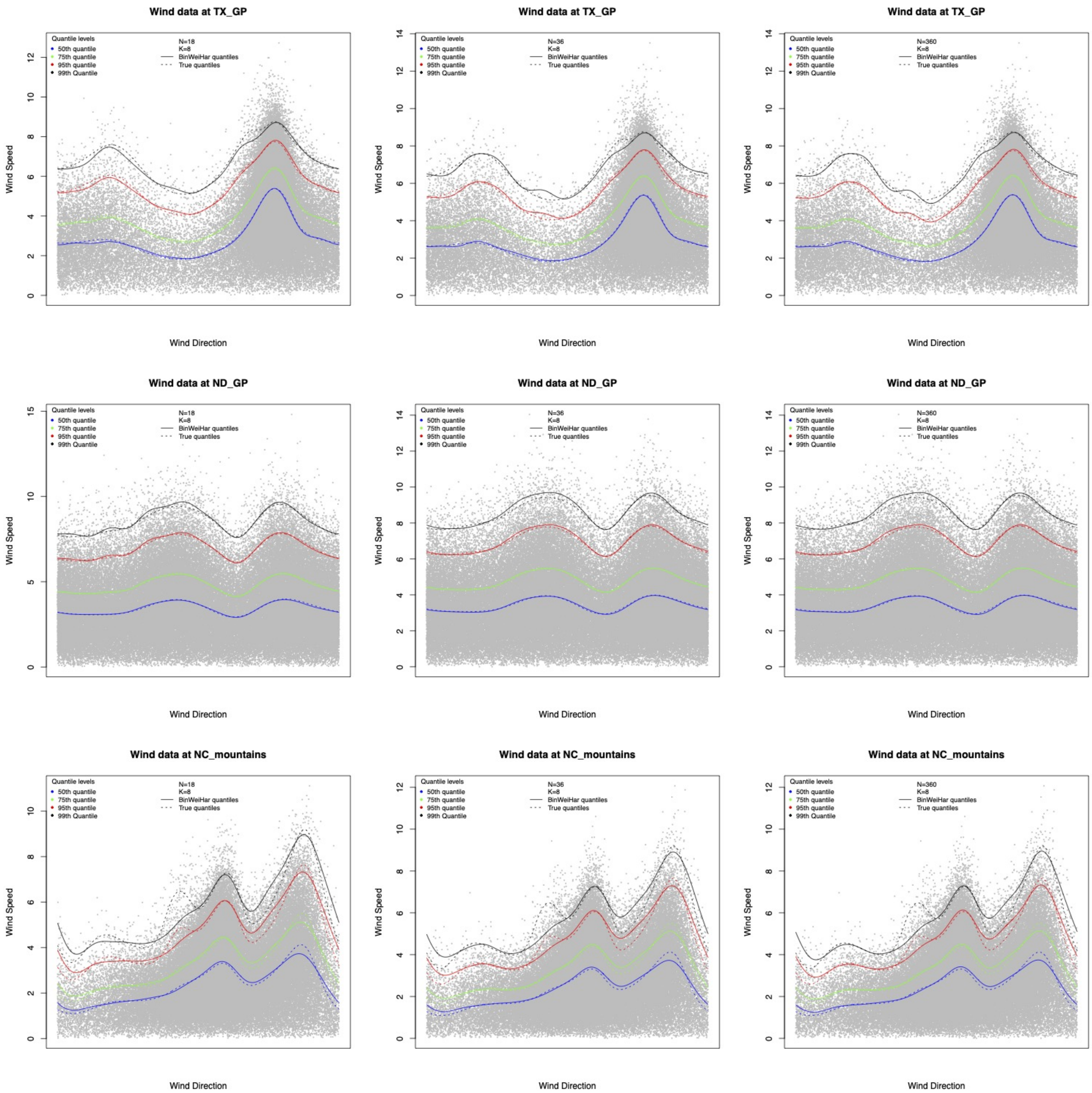

Figure 3: Qquantile curves of the directional wind speed distribution estimated using the BinWeiHar method when the sample size $n = 100000$ and $N = 18$ (left column), $N = 36$ (middle column), $N = 360$ (left columns). Each row represents a different location.

SM 3: Analysis on choosing the number of Fourier series

In this section we perform an analysis on the choice of the number of pairs of Fourier series, K_α and K_β. In doing so we use $N = 18$ and $N = 36$ bins and in each case we analyze five different situations: (i) K_α and K_β are both chosen using BIC, (ii) $K_\alpha = 4$ and K_β is chosen using using BIC, (iii) $K_\alpha = 8$ and K_β is chosen using using BIC, (iv) $K_\alpha = K_\beta = 4$,

(v) $K_\alpha = K_\beta = 8$. The fixed values of K_α and K_β are chosen such that the harmonic regression in Step 2 doesn't become underdetermined, i.e. we must have $N \geq 2K + 2$, where K represents K_α or K_β. We plot WIMRE for each of the five cases in Fig. 4 - Fig. 6.

We conclude this section by suggesting to use $K_\alpha = K_\beta = 8$ with $N = 36$ bins.

95% quantile BinWeiHar

	K_a, K_b from BIC	**K_a = 4, K_b from BIC**	**K_a = 8, K_b from BIC**	**K_a = K_b = 4**	**K_a = K_b = 8**
N = 18	**0.0111057405461189**	**0.0143681920625585**	**0.0111057405461189**	0.0151052293957332	**0.0112084742637167**
N = 36	0.0163102034248188	0.0168236441406529	0.0210556601856651	**0.0143046175867679**	0.0122377167251074

75% quantile BinWeiHar

	K_a, K_b from BIC	**K_a = 4, K_b from BIC**	**K_a = 8, K_b from BIC**	**K_a = K_b = 4**	**K_a = K_b = 8**
N = 18	**0.0133840201401709**	**0.0143833928998875**	**0.0133840201401709**	0.0191282493529481	**0.013316760257974**
N = 36	0.0184710706935268	0.019774937107861	0.0205630970715956	**0.0190401798894232**	0.0159884242667349

50% quantile BinWeiHar

	K_a, K_b from BIC	**K_a = 4, K_b from BIC**	**K_a = 8, K_b from BIC**	**K_a = K_b = 4**	**K_a = K_b = 8**
N = 18	0.0257743585387422	0.0254696248834977	0.0257743585387422	0.0276297977408933	0.0256166050777791
N = 36	**0.0213279442387626**	**0.0202043888120995**	**0.0205504205170147**	**0.0236621268955114**	**0.0193702477784595**

Figure 4: WIMRE values at Texas Great Plain (TX_GP) when K_α and K_β values are chosen using BIC (first column), K_α is fixed and K_β is chosen using BIC (second and third columns) and $K_\alpha = K_\beta$ are fixed (last two columns). Each table represents a different conditional quantile.

95% quantile BinWeiHar

	K_a, K_b from BIC	K_a = 4, K_b from BIC	K_a = 8, K_b from BIC	K_a = K_b = 4	K_a = K_b = 8
N = 18	**0.0205621832463836**	**0.019099306879011**	**0.0205621832463836**	0.0146796074610404	0.0205621832463836
N = 36	0.0253165508764861	0.0279329389324164	0.0269075494893134	**0.0136348390874929**	**0.0198510310896356**

75% quantile BinWeiHar

	K_a, K_b from BIC	K_a = 4, K_b from BIC	K_a = 8, K_b from BIC	K_a = K_b = 4	K_a = K_b = 8
N = 18	**0.0188532763005915**	**0.0187233121820366**	**0.0188532763005915**	**0.0125401809362793**	**0.0188532763005915**
N = 36	0.0254933718634748	0.0259272567762946	0.0258226033900824	0.0125624324969163	0.0191009964124994

50% quantile BinWeiHar

	K_a, K_b from BIC	K_a = 4, K_b from BIC	K_a = 8, K_b from BIC	K_a = K_b = 4	K_a = K_b = 8
N = 18	**0.0216836256122331**	**0.0215796825984641**	**0.0216836256122331**	0.016821834338898	**0.0216836256122331**
N = 36	0.0282671143398497	0.0280518219882675	0.0280697872821988	**0.0160582282594825**	0.0223069447337908

Figure 5: WIMRE values at North Dakota GP when K_α and K_β values are chosen using BIC (first column), K_α is fixed and K_β is chosen using BIC (second and third columns) and $K_\alpha = K_\beta$ are fixed (last two columns). Each table represents a different conditional quantile.

95% quantile BinWeiHar

	K_a, K_b from BIC	K_a = 4, K_b from BIC	K_a = 8, K_b from BIC	K_a = K_b = 4	K_a = K_b = 8
N = 18	0.0554739285694164	0.0515341615792185	0.0554739285694164	0.0624489651951732	0.0557380363671729
N = 36	**0.0512397117098671**	**0.0501693453338259**	**0.0509547271082063**	**0.0574833062375873**	**0.0484468021376172**

75% quantile BinWeiHar

	K_a, K_b from BIC	K_a = 4, K_b from BIC	K_a = 8, K_b from BIC	K_a = K_b = 4	K_a = K_b = 8
N = 18	**0.041541965992776**	**0.0410277999970727**	**0.041541965992776**	0.0498763790629277	0.0415973017671192
N = 36	0.0451700063999011	0.0465884217621606	0.0468538504243756	**0.0468238976653014**	**0.0404966690399305**

50% quantile BinWeiHar

	K_a, K_b from BIC	K_a = 4, K_b from BIC	K_a = 8, K_b from BIC	K_a = K_b = 4	K_a = K_b = 8
N = 18	**0.0482529347361654**	**0.0482281913704557**	**0.0482529347361654**	**0.0514493120791879**	**0.048295662382577**
N = 36	0.0568240685885806	0.0546545726739158	0.0549279679854316	0.0534650108411175	0.0505452257388911

Figure 6: WIMRE values at North Carolina mountains when K_α and K_β values are chosen using BIC (first column), K_α is fixed and K_β is chosen using BIC (second and third columns) and $K_\alpha = K_\beta$ are fixed (last two columns). Each table represents a different conditional quantile.

SM 4: Analysis of estimating the directional wind speed using BWHR with K=4 pairs of Fourier series vs. BWHR with pbs() and $df = 18$

We perform a study on estimating the directional wind speed distribution using `BWHR` (also refered to as BinWeiHar) method with $K_\alpha = K_\beta = 8$ pairs of Fourier series and `BWHR` where the parameter estimates are regressed on periodic B-splines with $Df = 18$. The WIMRE values are shown in Fig. 7 and they suggest that the `BWHR` method using $K_\alpha = K_\beta = 8$ outperforms the `BWHR` method using `pbs()` at all three locations.

	TX_GP	ND_GP	NC_mountain
BinWeiHar pbs	0.1126	0.2318	0.1101
BinWeiHar	0.0168	0.0210	0.0737

	TX_GP	ND_GP	NC_mountain
BinWeiHar pbs	0.0347	0.0807	0.0698
BinWeiHar	0.0143	0.0196	0.0695

	TX_GP	ND_GP	NC_mountain
BinWeiHar pbs	0.0364	0.0872	0.0956
BinWeiHar	0.0192	0.0231	0.0748

Figure 7: WIMRE values in estimating the conditional curves using the `BWHR` (refered to as BinWeiHar in the table) method where the parameter estimates are regressed on periodic B-splines (**first row of each table**) and $K_\alpha = K_\beta = 8$ pairs of Fourier series (**second row of each table**). **Top table** represents the 95% conditional quantile curve, **middle table** represents the 75% conditional quantile curve and **third table** represents the 50% conditional quantile curve.

SM 5: Sensitivity analysis on choosing the degrees of freedom (df) for BPQR

We estimate the directional wind speed distribution using `BPQR` with different df values and we show the plots of the estimated conditional quantile curves in Fig. 8. We observe that quantile curves estimated with lower df (5, 10) result in a smoothed curve while higher df value (30) leads to an overfitted curve. To balance the bias and variance we choose $df = 18$.

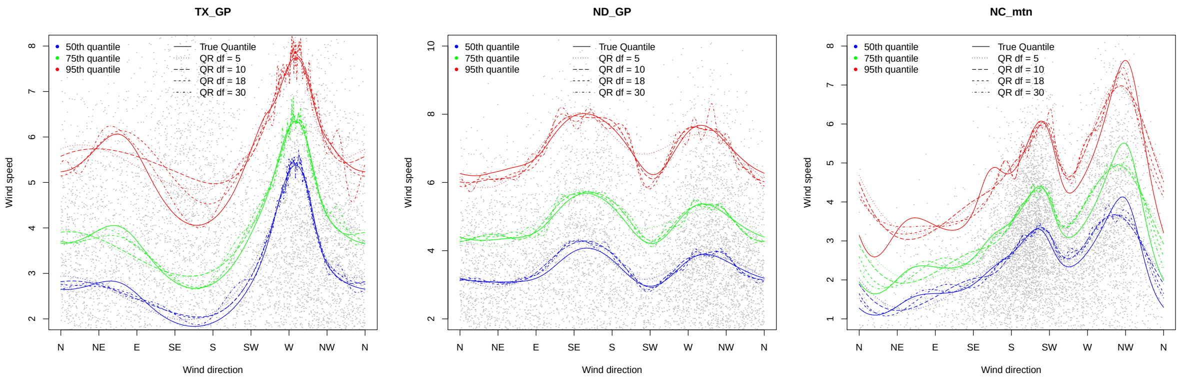

Figure 8: 95% conditional quantile (red), 75% conditional quantile (green) and 50% conditional quantile (blue) curves estimated with `BPQR` method using $df = 5$ (dotted line), $df = 10$ (long dashed line), $df = 18$ (dashed line) and $df = 30$ (dotted dashed line) along with the true quantile curves (solid line).

SM 6: Analysis using the mean square error (MSE) and the integrated mean square error (IMSE)

In our simulation study, we sate that it is important to incorporate the wind direction distribution when quantifying the estimation performance of the `BWHR` and `BPQR` methods. This conclusion is made based on our initial performance assessment using MSE. Here we evaluated the estimation performance of the two methods using the MSE between the true and estimated conditional quantiles, denoted by q and $\hat{q}$, respectively, over N Monte Carlo replicates ($N = 500$ in our stdy). The MSE at a given quantile level, $\tau \in [0, 1]$, and angle, $\phi \in [0, 2\pi]$ is computed at a set of m evenly spaced wind direction grid points $\{\phi_1, \phi_2, \ldots, \phi_m\}$ with a sufficiently large m ($m = 629$ in our study). Specifically, for each angle ϕ_i, $i = 1, \ldots m$, we have:

$$\mathrm{MSE}_{\hat{q}_{\phi_i}} = \frac{1}{N} \sum_{j=1}^{N} (\hat{q}_j(\phi_i) - q(\phi_i))^2, \tag{1}$$

Furthermore, the MSE_{ϕ_i} values are averaged over the discretized wind direction domain using the following formula:

$$MSE = \frac{1}{m} \sum_{i=1}^{m} MSE_{\phi_i}$$

The $\mathrm{MSE}_{\hat{q}_{\phi_i}}$, $i = 1, \ldots, m$, and MSE values considering the 95% quantile level are depicted in the top row of Fig. 9. A quick analysis of the graphs should reflect that for most directions the `BWHR` method has a smaller MSE than `BPQR`. However, there are wind directions for which the MSE of the `BWHR` method is higher than that of `BPQR`. Comparing these results to the estimated wind direction distribution, depicted in the middle row of Fig. 9, we see that these circumstances happen when the density of wind direction is very small. This suggests the inclusion of the density of wind direction when computing the error. Hence, the use of the integrated square error defined as follows:

$$\text{IMSE} = \int_0^{2\pi} MSE_\phi \cdot f_\Phi(\phi)\, d\phi,$$

where $f_\Phi(\phi)$ represents the density of wind direction (ϕ). To compute this value we proceed similarly to the MSE case, where we dicretize the wind direction domain to a set of grid points $\{\phi_1, \phi_2, \ldots, \phi_m\}$ and the IMSE_ϕ along with the IMSE is computed using the following formulas:

$$\begin{aligned} \text{IMSE}_{\phi_i} &= \frac{1}{500} \sum_{j=1}^{500}, f_{\Phi_i}(\phi_i)(q_{est,j}(\phi_i) - q_{true}(\phi_i))^2, \quad i = 1, \ldots m, \\ \text{IMSE} &= \frac{\sum_{i=1}^m \text{IMSE}_{\phi_i}}{\sum_{i=1}^m f_{\Phi_i}(\phi_i)}. \end{aligned}$$

The results of the IMSE_{ϕ_i}, $i = 1, \ldots n$, along with the IMSE values for the 95% quantile level are depicted in the bottom row of Fig. 9 (Fig. 10 for quantile differences) showing a decreased error in the `BWHR` model. Furthermore, in Fig. 11 we plot the distribution of the IMSE for the 95% quantiles of the `BWHR` and `BPQR` models. As the boxplots show, the distribution of IMSE of the `BWHR` method is less variable than that of `BPQR`. Finally, the scatter plots included in the bottom row of Fig. 11 tell us that `BWHR` model has a lower IMSE than `BPQR` 76.6% of times at the TX_GP location, 94% of times at ND_GP and 95.5% of times at NC_mountains.

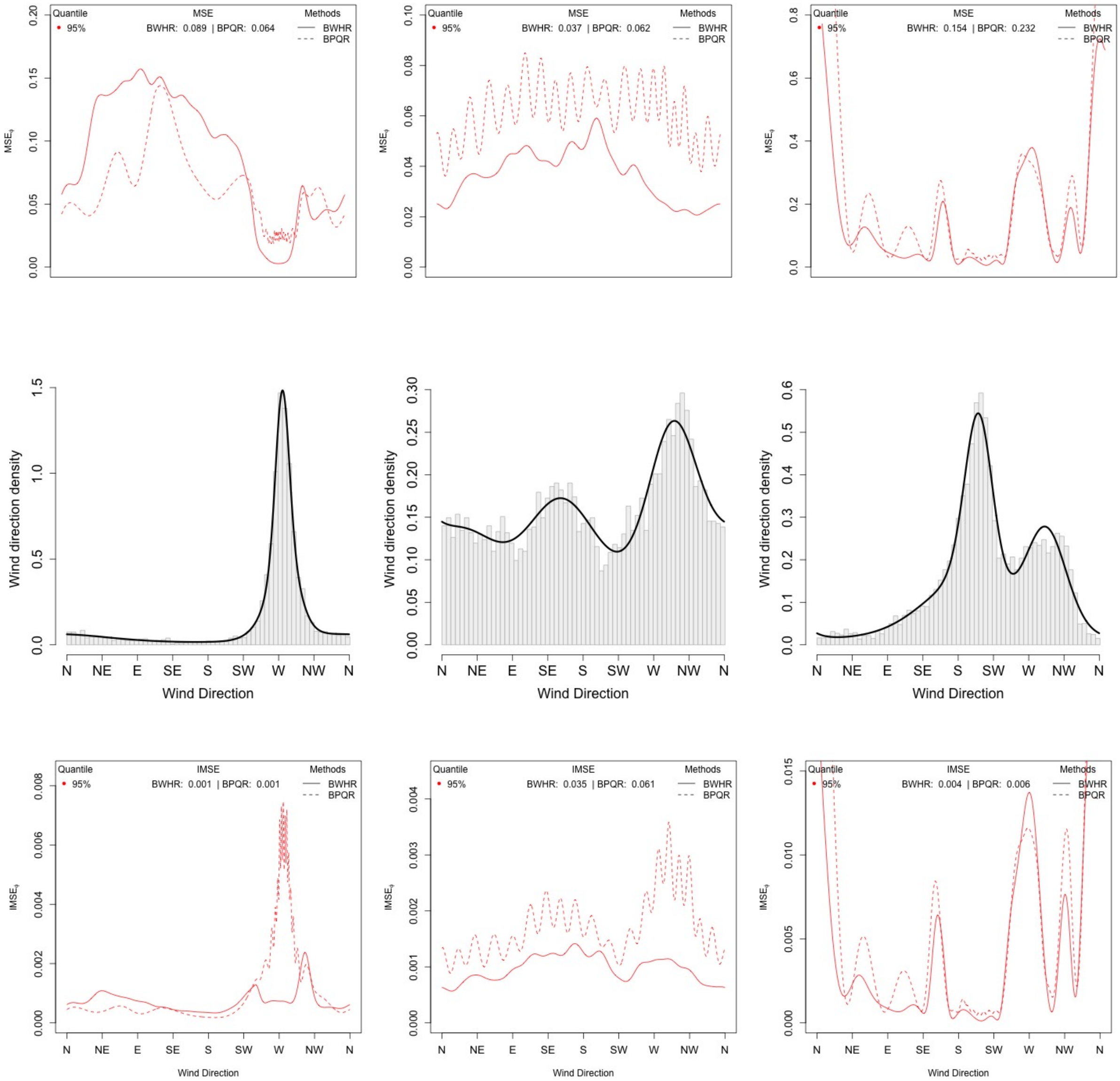

Figure 9: **Top row**: MSE_ϕ and MSE values for the 95% conditional quantile of the `BWHR` (solid line) and of the `BPQR` method (dashed line) at three different locations; **middle row:** estimates of the wind direction distribution at three different locations; **bottom row:** $\mathrm{IMSE}_\phi$and IMSE values for the 95% conditional quantile of the `BWHR` method (solid line) and of the `BPQR` method (dashed line) at three different locations.

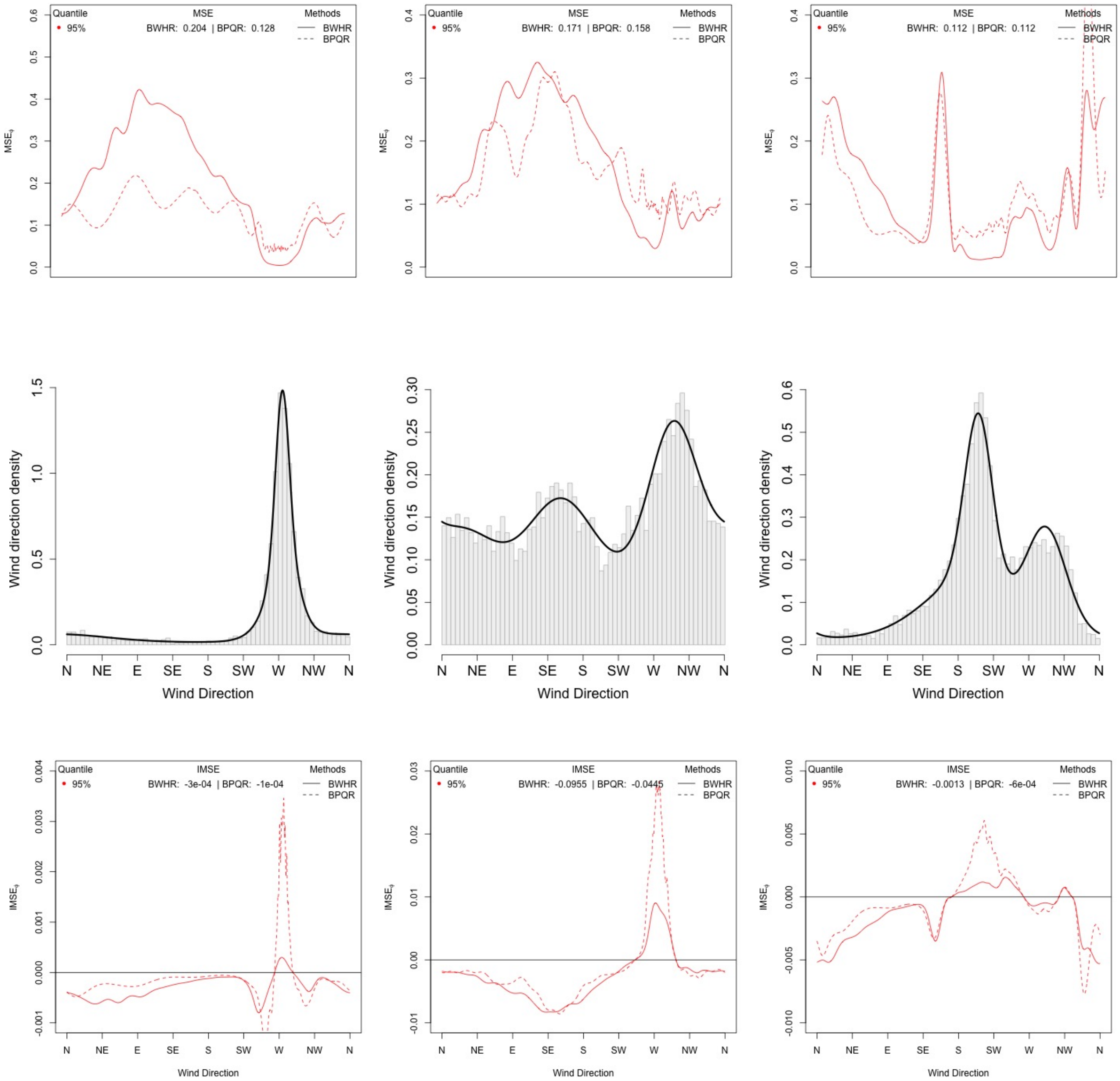

Figure 10: As in Fig 9 but for quantile differences.

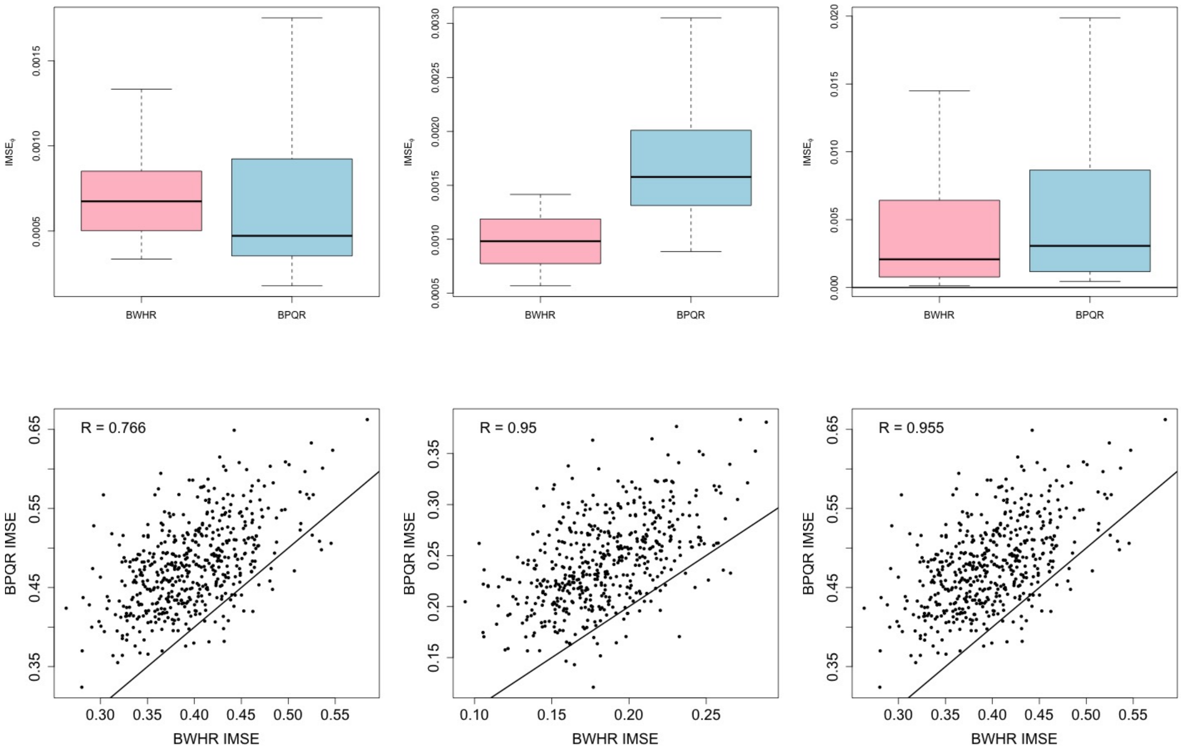

Figure 11: **Top row:** Boxplots of the distribution of the IMSE$_\phi$ of the 95% quantile curves; **bottom row:** Scatter plots of the `BWHR` vs the `BPQR` IMSE$_\phi$ where the solid line represents the line $y = x$ and R is the ratio of number of time the `BWHR` method has a lower IMSE compared to `BPQR`. Each column represents a different location.

The von Mises mixture distribution is evaluated using IMSE$_\phi$ and IMSE defined above. Specifically,

$$\text{IMSE}_{\phi_i} = \frac{1}{500}\sum_{j=1}^{500} f_{true}(\phi_i)(f_{est,j}(\phi_i) - f_{true}(\phi_i))^2, \quad i = 1, \ldots m,$$
$$\text{IMSE} = \frac{\sum_{i=1}^{m} \text{IMSE}_{\phi_i}}{\sum_{i=1}^{n} f_{true}(\phi_i)} \ ,$$

where $f_{true}(\phi_i)$ and $f_{est,j}(\phi_i)$, $j = 1, \ldots, 500$, denote the true and estimated wind direction distribution at a given angle ϕ_i, $i = 1, \ldots, n$. The results are plotted in Fig. 12 and show a small IMSE value suggesting that a von Mises mixture distribution is a good fit to wind direction.

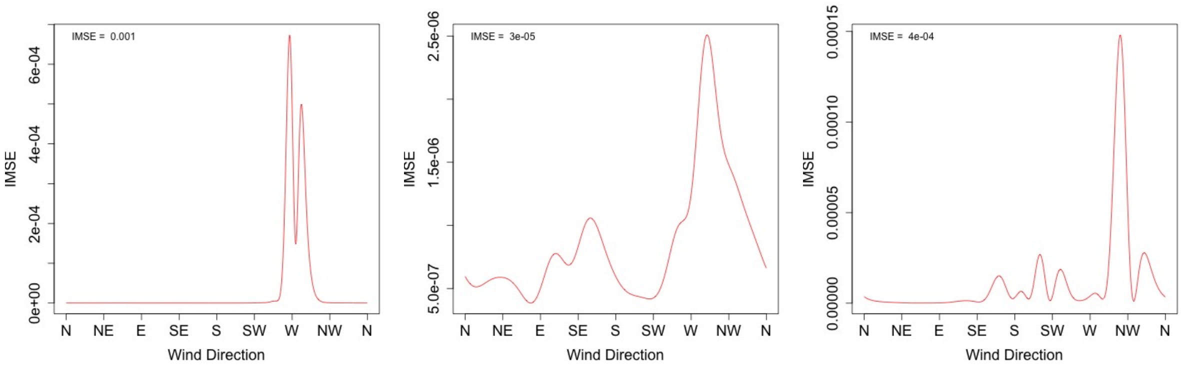

Figure 12: IMSE_ϕ values for a set of discretized wind directions along with the IMSE values at each of the three locations.