

Statistical Approaches for Wind Resource Assessment

Kalyan Veeramachaneni, Xiang Ye, and Una-May O'Reilly

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology, USA

Abstract. We describe and evaluate a set of computational intelligence techniques for long-term wind resource assessment. Short term sensor measurements at a potential wind farm site are correlated with publicly available wind data sources in close spatial proximity in order to extrapolate long-term predictions for the site. This general approach to assessment is called “MCP”: Measure, Correlate and Predict. Our techniques are based upon statistical inference. They aim to address accurately correlating inexpensive but noisy, short term measurements at the site. Each technique relies upon estimating the joint distribution of wind speeds at the site and the publicly available neighbouring sources. For a site at the Boston Museum of Science when the availability of site data varies between 3, 6 and 8 months, we find that copula modeling is robust to data availability and consistently best overall.

Keywords: wind resource assessment, Measure-Correlate-Predict, multivariate modeling, graphical modeling, copula estimation

1 Introduction

A wind resource assessment is initiated as part of the decision process leading to selecting a wind farm site. There are multiple factors which influence site selection. Among them are legal considerations, community opinion, ease of construction, maintenance and cabling cost. Arguably most crucial is whether there is enough wind in the ideal speed range that will endure over a long span of times such as a generation or longer. Prediction of wind at high frequency like hours to days to weeks is fraught with technical and sensing challenges plus intrinsic uncertainty. Wind resource assessment for site selection contrasts with high frequency prediction. Its goal is to provide a general estimate that guides selection without being a precise prediction. The annual, actual wind resource of a farm would be expected to deviate from the assessment with reasonable variance. However, when the actual annual resource is averaged over a long span, the goal is that the assessment and actuality should match up. In this way wind resource assessment helps inform the question of the production capacity of the site over its extended lifetime (which potentially includes successive upgrades of turbines and related facilities).

A wind resource assessment is presented as a set of probability distributions of wind speed for directional intervals which span 360° . An example of 3 distributions, for the intervals $0^\circ - 15^\circ$, $15^\circ - 30^\circ$, and $30^\circ - 45^\circ$ is shown on the left of Figure 1. Each plotted probability function is modeled with a Weibull distribution which is parameterized by shape and scale. Integrating this function (mathematically) allows one to derive the probability that the wind speed from a given direction range will be within a specific range.

The assessment can also be visualized via a wind rose, see Figure 1, right. The span of the entire 360 degrees is oriented in North-South compass direction to inform its alignment to the site. Figure 1, left shows 16 direction intervals, each as a discrete “slice” with coloring that depicts wind speed. The length and width of the slice conveys probability.

There are multiple methodologies which derive a wind resource assessment. All are subject to great uncertainty. When a wind resource assessment is based upon wind maps and publicly available datasets from the closest locations, it tends to overestimate the wind speed because the maps are so macroscopic. Even when the resource estimated by the wind map for a geographical location is improved upon by utilizing a model that accounts for surface roughness and other factors, significant inaccuracies persist because specific details of the site remain neglected. Alternatively, a computational fluid dynamics (CFD) model can be used to achieve a better resource assessment. CFD, too, however has limitations. It is very hard to incorporate all the local attributes and factors related to turbulence into the simulation. While the wind industry has started to combine CFD and wind map approaches, the current methods are ad-hoc, not robust and more expensive than desired.

In this chapter we provide new techniques for the only assessment methodology that takes into account as many years of historical data as possible (though that data is remote from the site itself), while also integrating site specific information, albeit short term and relatively noisy. We consider the Measure-Correlate-Predict assessment methodology, abbreviated as MCP, which exploits anemometers, and/or other sensing equipment that provide site specific data [1–4]. The Measure step involves measuring wind speed and direction at the site for a certain duration of time. In the Correlate step, this data is then associated with simultaneous data from nearby meteorological stations, so-called **historical sites** which also have long term historical data. A correlation model is built between the time-synchronized datasets. In the Predict step, the model is then used along with the historical data from the meteorological stations to predict the wind resource at the site. The prediction is expressed as a bivariate (speed and direction) statistical distribution or a “wind rose” as shown in Figure 1.

While MCP does incorporate site-specific data, the data is based upon very inexpensive sensors, i.e anemometers, which are consequently very noisy. Additionally, anemometers are frequently moved on the site and not deployed for any significant length of time. Thus, the key challenge of MCP is to accurately predict with cheap, noisy, short-term sensor measurements. We aim to address how techniques match up with data availability. Usually 8-10 months are consid-

ered as a standard in the wind industry. However, we seek very simple modeling techniques which can reduce the amount of data required to get an accurate estimation of the long term wind speed.

Another challenge is integrating historic site data. The best historic site (e.g. airport) to correlate wind speed with a site might be somewhat intuitive if only wind direction is considered: for example, when the wind blows from the north at the site, its speed might be best correlated with speeds from a particular airport meteorological tower immediately to the north. However, site conditions such as local terrain height variation or terrain features such as forests and large buildings make identifying the best correlation sources much more complicated. The sites' conditions cannot be observed directly in the data but influence the correlative relationships. Also, the strength of correlations may be sensitive to the how wind direction is segmented into aggregative bins and intervals. To date, the selection is done with trial and error modeling and ad hoc understanding of the similarity between the proposed and historic sites. Almost always, a single historic site, closest to the test site, is used to build the predictive model. Yet, for any directional bin, better accuracy is obtainable when multiple historic sites in different directions with respect to the site are candidates for integration into the estimation. In this chapter we address how to integrate the wind information from multiple historic sites and weigh each historic site according to its correlative strength.

It is not uncommon for wind farms to fall short of their expected energy output. A study performed in the United Kingdom monitored small and micro wind turbine performance and found that the wind resource was over-predicted by a factor of 15 [5]. In another example, data analysis of 19 small turbines installed in Massachusetts [6], showed, on average, turbines underperforming by a factor of 4. The capacity factor was found to be as low as 3%-4%. Inadequate wind resource assessment was determined to be one of the major factors (20%) influencing the underperformance.

In this chapter we describe a set of computational intelligence techniques based upon statistical inference. Each technique relies upon estimating the joint distribution of wind speeds at the site and the publicly available neighboring sources. The accuracy of any one of these techniques is sensitive to choices in the modeling setup and/or parameterizations. We aim to assess which technique and choices are best. Our assessment will utilize anemometer measurements for a site at the Boston Museum of Science when the availability of site data varies between 3, 6 and 8 months while correlating with data from 14 airports nearby, see Figure 4 and Table ??.

We proceed as follows: Section 2 presents a detailed description of MCP. Section 3 presents statistical techniques which can be used in an MCP framework. Section 4 presents the means by which we evaluate the techniques. Section 5 presents the empirical evaluation. Finally Section 6 states our conclusions and outlines future work.

2 Measure-Corelate-Predict (MCP)

We consider wind resource estimation derived by a methodology known as Measure-Correlate-Predict or MCP, see Figure 2. In terms of notation, the wind at a particular location is characterized by speed denoted by x and direction θ . Wind speed is measured by anemometers and wind direction is measured by wind vanes. The 360° direction is split into multiple bins with a lower limit (θ_l) and upper limit (θ_u). We give an index value of $J = 1 \dots j$ for the directional bin. We represent the wind speed measurement at the test site (where wind resource needs to be estimated) with y and the other sites (for whom the long term wind resource is available) as x and index these other sites with $M = 1 \dots m$.

The three steps of MCP are:

MEASURE Short term sensing measurements on the site are collected. This is denoted by $Y = \{y_{t_k} \dots y_{t_n}\}$. Measurements can be collected using anemometers on the site, a newly-constructed meteorological tower, or even remote sensing technologies such as sonar or lidar. Different measurement techniques incur different costs that dictate their feasibility for different projects. Measurements from nearby sites for the same period are gathered. These sites, called *historical sites*, have additional data for the past 10–20 years. These are denoted by $X = \{x_{t_k \dots t_n}^{1 \dots m}\}$ where each $x_{t_k \dots t_n}^i$ corresponds to data from one historical site, k and n are time indices, and m denotes the total number of historical sites. Historical data that is not simultaneous in time to the site observations used in modeling will be used in the PREDICT step.

CORRELATE A single directional model is first built correlating the wind directions observed at the site with simultaneous historical site wind directions. Next, for each directional interval, called a (directional) bin, of a 360° radius, a model is built correlating the wind speeds at the site with simultaneous speeds at the historical sites, i.e. $Y_{t_i} = f_{\theta_j}(x_{t_i}^{1 \dots m})$ where $k \leq i \leq n$. The data available from the site at this stage is expected to be sparse and noisy.

PREDICT To obtain an accurate estimation of long term wind conditions at the site, we first divide the data from the historic sites (which is not simultaneous in time to the site observations used in modeling) into subsets that correspond to a directional bin. Prediction of the long term site conditions follows two steps:

- A** : We use the model we developed for that direction f_{θ_j} and the data from the historic sites corresponding to this direction $x_{t_1 \dots t_{k-1}}^{1 \dots m} | \theta_j$ to predict what the wind speed $\mathbf{Y}_p = y_{t_1 \dots t_{k-1}}$ at the site would be.
- B** : With the predictions \mathbf{Y}_p , from **A** above, we estimate parameters for a Weibull distribution. This distribution is our answer to the wind resource assessment problem. We generate a distribution for each directional bin. A few example distributions for different bins are shown in Figure 1, left. Alternatively, these distributions can be summarized via a wind rose also shown in Figure 1, right.

The goal is to generate a predicted long term wind speed distribution in each direction which will be as close as possible to the real (as yet unexperienced)

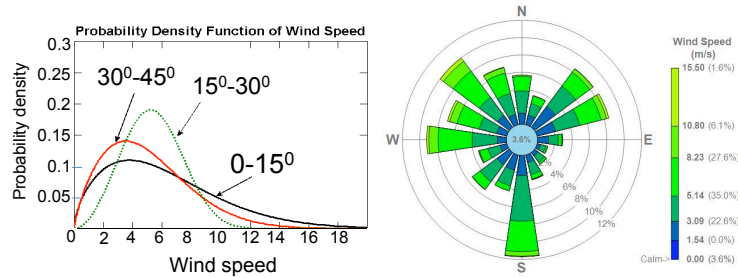


Fig. 1. A wind resource estimation is expressed as a bivariate (speed and direction) statistical distribution or a “wind rose”.

distribution. The result from MCP, i.e. the statistical distribution in each bin, is then used to estimate the energy which can be expected from a wind turbine, given the power curve supplied by its manufacturer. This calculation can be extended over an entire farm if wake interactions among the turbines are taken into account. See [7] for more details.

Note that distribution not only captures the mean, but also variance in this speed. This is critical for assessment of long term wind resource and the long term energy estimate.

A variety of methods are developed in [8] to evaluate the accuracy of the predicted wind speed distribution. One method measures the accuracy in terms of ratios between true and actual parameters of the Weibull distribution. That is, true shape versus estimated shape and true scale versus estimated scale. To completely capture any possible inaccuracy in the predicted distribution, we measure a symmetric Kullback-Leibler distance. It is important to note that this measure is different than the mean-squared error or mean-absolute error which measure the accuracy in terms of difference between each predicted value and the true observation. Methods that minimize these errors would not necessarily accurately express how close the approximation is to the true distribution.

We now proceed to describe the set of statistical approaches we introduce for deployment within the MCP framework.

3 Methodology for wind speed estimation

Notationally, we refer to a training point as $l \in \{1 \dots L\}$ and a point for which we have to make prediction as $k \in \{1 \dots K\}$. We drop the notation for time after having time synchronized all the measurements across locations. We also drop the subscript for directional bin. From this point onwards when we refer to a model, it is the model for a particular bin j . $f_Z(z)$ refers to a probability density function of the variable (or set of variables) z . $F_Z(z)$ refers to cumulative distribution function for the variable z such that $F_Z(z = \alpha) = \int_{-\inf}^{\alpha} f_Z(z)$ for a continuous density function.

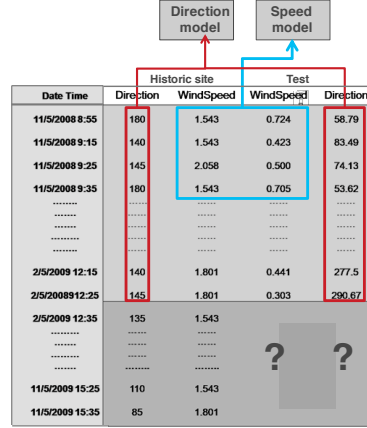


Fig. 2. MCP generates a model correlating site wind directions to those simultaneously at historical sites. For a directional bin, it generates a model correlating simultaneous speeds.

Our methodology for MCP has four steps.

- Step 1:** To start, we build a multivariate distribution with the probability density function $f_{\mathbf{X},Y}(\mathbf{x}, y)$, where $\mathbf{x} = \{x_1 \dots x_m\}$ are the wind speeds at the historic sites and y is the wind speed at the site. To do this we employ likelihood parameter estimation. The model building process is similar for all the bins and only the data needs to be changed.
- Step 2:** Given the joint distribution from Step 1, we predict the probability density of y that corresponds to a given test sample $\mathbf{x}_k = \{x_{1_k} \dots x_{m_k}\}$ by estimating the conditional density $f_Y(y|\mathbf{x}_k)$. The conditional can be estimated by:

$$f_{Y|\mathbf{X}=\mathbf{x}_k}(y|\mathbf{x}_k) = \frac{f_{\mathbf{X},Y}(\mathbf{x}_k, y)}{\int_y f_{\mathbf{X},Y}(\mathbf{x}_k, y) dy}. \quad (1)$$

- Step 3:** We now can make a point prediction of \hat{y}_k by finding the value for y that maximizes the conditional.

$$\hat{y} = \arg \max_{y \in Y} f(y|\mathbf{X} = \mathbf{x}_k). \quad (2)$$

- Step 4:** All the predictions for $\hat{y}_{1 \dots K}$ are estimated to a Weibull density function which gives an estimate of the long term wind resource at the test site.

The methodology implies two design decisions:

- Choosing a model** A key decision is which density function should be used to model the univariate densities $f_{X_i}(x_i)$ and the choice of the joint density

function. A simple and straight forward choice is univariate and multivariate Gaussians. This is because Gaussian density functions have closed form analytic equations for estimation of parameters and evaluating conditionals that are readily computed. Unfortunately, in many cases this choice could lead to inaccuracies. Individual variables may not fit a Gaussian density function without significant error, e.g. when they have significant tail properties or bi-modal distribution. A joint Gaussian density function only captures the linear correlation between the variables. If we choose non-Gaussian univariate densities for x_i , we have to employ *copulas* to construct a multivariate joint density function.

Assumptions Regarding Variable Dependency Structure Parametric estimation of a joint density function is expensive to compute and requires a large amount of data. Inference from the joint is also expensive. Assuming conditional independency among some variables offers efficiency in all three respects because it allows the density function to be factored. As an example, Figure 3 shows two different possible dependency structures for the variables and the output variable y . On the right is the simplest structure which assumes independence among the input variables given the output variable. The plot on the left shows a possible structure that could be learnt from the data.

In wind resource assessment a primary advantage of learning the dependency structure is the reduction in prediction time despite incorporating more airports into a model while not requiring more site data. Herein, we will evaluate whether the the structure should be learned from the data or predefined.

Each of our statistical techniques for wind resource assessment follow these steps. Each is distinctive in terms of Step 1 and how it estimates the joint multivariate density function. We now proceed to describe each of them.

3.1 Multivariate normal model

Here a multivariate normal model $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is assumed for the joint density $f_{\mathbf{X}, Y}(\mathbf{x}, y)$ given by

$$f_Z(\mathbf{z}) = (2\pi)^{-(m+1)/2} \det(\boldsymbol{\Sigma})^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu})\right) \quad (3)$$

where $z = \{\mathbf{x}, y\}$ and *det* represents determinant of a matrix. The multivariate Gaussian density function has two parameters $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$. We estimate these parameters by maximizing the likelihood function given a set of L *i.i.d* observations. The greatest advantage of this model is the ease with which the model can be built since the maximum likelihood estimates (MLE) simply are given by closed form analytic forms. The MLE for the mean vector $\boldsymbol{\mu}$ for the variates \mathbf{z} is simply the sample mean. The MLE of the covariance matrix $\boldsymbol{\Sigma}$ is given by:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^n (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^T \quad (4)$$

Once we estimate the parameters for the joint density given the training data, using closed form expressions, we use the joint density function to derive the conditional density for y given \mathbf{x}_k sample in the testing data. This density is also Gaussian and has a mean $\mu_{y|\mathbf{x}_k}$ and variance $\sigma_{y|\mathbf{x}_k}$. The value $\mu_{y|\mathbf{x}_k}$ is used as the point prediction for \hat{y}_k for the given \mathbf{x}_k . The variance $\sigma_{y|\mathbf{x}_k}$ provides the uncertainty around the prediction. If $\sigma_{y|\mathbf{x}_k}$ is high, the uncertainty is high.

3.2 Non-parametric multivariate model

Our second model is adapted from [9] in which the authors employ a multivariate kernel density estimator. A Gaussian kernel is chosen and the density is estimated using all the data $\mathbf{x}_{1\dots L}$. The multivariate kernel density function is given by

$$f_{\mathbf{X},Y}(\mathbf{x}, y) = \sum_{l=1}^L \prod_{j=1}^m K(x_j - x_{j,l}) K(y - y_l) \quad (5)$$

For a test point \mathbf{x}_k , for which we do not know the output, a prediction is made by finding the expected value of the conditional density function $f_Y(y|\mathbf{x}_k)$ given by

$$E(Y|\mathbf{X} = \mathbf{x}_k) = \int_y y \cdot f_Y(y|\mathbf{x}_k) dy \quad (6)$$

$$= \int_y y \cdot \frac{\frac{1}{L} \sum_{i=1}^L \prod_{j=1}^m K(x_{j,k} - x_{j,i}) K(y - y_i)}{\frac{1}{L} \sum_{i=1}^L \prod_{j=1}^m K(x_{j,k} - x_{j,i})} dy \quad (7)$$

$$= \frac{\sum_{i=1}^L y_i \cdot \prod_{j=1}^m K(x_{j,k} - x_{j,i})}{\sum_{i=1}^L \prod_{j=1}^m K(x_{j,k} - x_{j,i})}. \quad (8)$$

In this model there are no parameters and there is no step for estimation. Although eq. (5) presents the density function it is not evaluated unless we see a new testing point. For us to be able to evaluate the expected value of the wind speed at the test site, we need to store all the training points and use them in eq. (8) to make predictions. In (8), given a test point $\mathbf{x}_k = \{x_{1,k} \dots x_{m,k}\}$, the kernel value is evaluated for difference between a training point \mathbf{x}_1 and this test point. This value is multiplied with the corresponding value of y_l . This is repeated for all the training points $1 \dots L$ and summed. This forms the numerator. The summation when done without the multiplication of y_l forms the denominator in eq. (8). This approach has a few drawbacks. The designer has to choose the kernel. Then the parameters for the kernel need to be tuned via further splitting of the training data. It also requires retaining all the training data in order to make predictions. The evaluation of the kernel is done for L times for each test point.

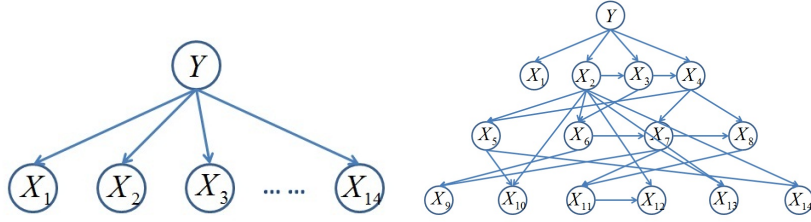


Fig. 3. Left: Naive Bayes structure. This structure is assumed and no learning is required. $x_1 \dots x_{14}$ represent the 14 variables from the airports and y represents the variable at the test site in Boston. Right: Structure is learnt using the K2 algorithm. A maximum of two parents is specified. x_2 emerges as a parent for most of the nodes.

3.3 Graphical model with naive structure

This technique uses a multivariate Gaussian model assuming independent, variable dependency structure. We model the joint density function as a Bayesian network with independent variables $x_{i \dots m}$, that is

$$f_{\mathbf{X}, Y}(\mathbf{x}, y) = \prod_{i=1}^m f_{X_i}(x_i | y) f_Y(y). \quad (9)$$

A Bayesian network $\mathcal{B} = \langle \mathcal{G}, \theta \rangle$ is a probabilistic graphical model that represents a joint probability distribution over a set of random variables X_1, \dots, X_n . The Bayesian network representation has two components. A directed acyclic graph (DAG) \mathcal{G} encodes independence relations between variables. Each variable is a node in this graph. A set of local probability models θ defines the conditional probability distribution of each node given its parents in the graph. Let Pa_X denote the parents of node X in \mathcal{G} . Then the network \mathcal{B} encodes the following probability distribution:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Pa}_{X_i}).$$

In comparison the multivariate model previously introduced assumes the same naive Bayes structure but uses kernel density estimators for each individual variable.

Estimation: Given this structure, also shown in the Figure 3, we estimate the parameters at each node in Figure 3(left) for the bivariate Gaussian density function $f_{X_i, Y}(x_i, y)$. We use maximum likelihood estimation and employ similar analytical expressions employed in Section 3.1 but this time for each pair x_i, y individually. Parameters for $f_Y(y)$ are estimated directly by using the training data for y .

Prediction: For a test point \mathbf{x}_k , we derive the conditional:

$$f_{Y|\mathbf{X}=\mathbf{x}_k}(y|\mathbf{x}_k) = \frac{\prod_{i=1}^m f_{X_i}(x_i|y)f_Y(y)}{\int_y \prod_{i=1}^m f_{X_i}(x_i|y)f_Y(y)dy}. \quad (10)$$

If we were to minimize the mean squared error, the optimal prediction is the mean of the conditional density function above. If we were to minimize the mean absolute error, the optimal prediction is the median of the conditional density function [10]. Since we assume a Gaussian model for the joint density function, this results in a Gaussian model for the conditional as well, the mean and median for $y|\mathbf{x}_k$ are the same. This is also the value that maximizes the conditional. Hence \hat{y}_k is

$$\hat{y}_k = \int_y y f_{Y|\mathbf{X}=\mathbf{x}_k}(y|\mathbf{x}_k) dy \quad (11)$$

3.4 Graphical model with structure learning

In the model of Section 3.3 we assume a naive structure. By contrast, in this technique, we learn the variable dependency structure (i.e. Bayesian network) given the site data. Many structure learning algorithms are available in literature and can be readily employed [11–13].

Estimation: Our estimation has two steps. First we learn the network structure \mathcal{G} and then estimate parameters for the conditional probability distribution at each node in \mathcal{G} . In our experiments we employ a heuristic called K2 [13]. The algorithm takes the order of the variables and attempts to learn a dependence structure. [11]. We then estimate the parameters at each node for the conditional joint density via likelihood parameter estimation.

Prediction: Given a new test point multiple inference techniques are available to predict the value of $y_k|\mathbf{x}_k$ [11].

We refer the reader to [11] for a thorough introduction to Bayesian networks and a well known publicly available MATLAB based tool for K2, likelihood parameter estimation and inference [14].

3.5 Multivariate copulas

Our previous modeling techniques assume a Gaussian distribution for all variables and a Gaussian joint for the multivariate. It is arguable however that Gaussian distributions do not accurately represent the wind speed distributions. In fact, conventionally a univariate Weibull distribution [15] is used to parametrically describe wind sensor measurements. A Weibull distribution is likely also chosen for its flexibility because it can express any one of multiple distributions, including Rayleigh or Gaussian.

To the best of our knowledge, however, joint density functions for non-Gaussian distributions have not been estimated for wind resource assessment.

In this chapter, to build a multivariate model from marginal distributions which are not all Gaussian, we exploit *copula* functions. A *copula* framework provides a means of modeling a multivariate joint distribution from training data. It is then possible to proceed with inference from the copula function.

Because copula estimation is less well known, we now briefly review copula theory. We will then describe how we construct the individual parametric distributions which are components of a copula and then, how we couple them to form a multivariate density function. Finally, we present our approach to predict the value of y given $x_{1\dots m}$.

A copula function $C(u_1, \dots, u_{m+1}; \theta)$ with parameter θ represents a joint distribution function for multiple *uniform* random variables $U_1 \dots U_{m+1}$ such that

$$C(u_1, \dots, u_{m+1}; \theta) = F(U_1 \leq u_1, \dots, U_{m+1} \leq u_{m+1}). \quad (12)$$

Let $U_1 \dots U_m$ represent the cumulative distribution functions (CDF) for variables x_1, \dots, x_m and U_{m+1} represent the CDF for y . Hence the *copula* represents the joint distribution function of $C(F(x_1) \dots F(x_m), F(y))$, where $U_i = F(x_i)$. According to Sklar's theorem any *copula* function taking marginal distributions $F(x_i)$ as its arguments, defines a valid joint distribution with marginals $F(x_i)$. Thus we are able to construct the joint distribution function for $x_1 \dots x_m, y$ given by

$$F(x_1 \dots x_m, y) = C(F(x_1) \dots F(x_m), F(y); \theta) \quad (13)$$

The joint probability density function (PDF) is obtained by taking the $m + 1^{th}$ order derivative of the eq. (13)

$$f(x_1 \dots x_m, y) = \frac{\partial^{m+1}}{\partial x_1 \dots \partial x_m \partial y} C(F(x_1) \dots F(x_m), F(y); \theta) \quad (14)$$

$$= \prod_{i=1}^m f(x_i) f(y) c(F(x_1) \dots F(x_m), F(y)) \quad (15)$$

where $c(\cdot)$ is the *copula* density. Thus the joint density function is a weighted version of independent density functions, where the weight is derived via *copula* density. Multiple copulas exist in literature. In this chapter we consider a multivariate Gaussian copula to form a statistical model for our variables given by

$$C_G(\Sigma) = F_G(F^{-1}(u_1) \dots F^{-1}(u_m), F^{-1}(u_y), \Sigma) \quad (16)$$

where F_G is the CDF of multivariate normal with zero mean vector and Σ as covariance and F^{-1} is the inverse of the standard normal.

Estimation of parameters: There are two sets of parameters to estimate. The first set of parameters for the multivariate Gaussian copula is Σ . The second set, denoted by $\Psi = \{\psi, \psi_y\}$ are the parameters for the marginals of \mathbf{x}, y . Given N *i.i.d* observations of the variables \mathbf{x}, y , the log-likelihood function is:

$$L(\mathbf{x}, y; \Sigma, \Psi) = \sum_{l=1}^N \log f(\mathbf{x}_l, y_l | \Sigma, \Psi) \quad (17)$$

$$= \sum_{l=1}^N \log \left\{ \left(\prod_{i=1}^m f(x_{il}; \psi_i) f(y_l; \psi_y) \right) c(F(x_1) \dots F(x_m), F(y); \Sigma) \right\} \quad (18)$$

Parameters Ψ are estimated via [16]

$$\hat{\Psi} = \arg \max_{\Psi \in \psi} \left\{ \sum_{l=1}^N \log \left\{ \left(\prod_{i=1}^m f(x_{il}; \psi_i) f(y_l; \psi_y) \right) c(F(x_1) \dots F(x_m), F(y); \Sigma) \right\} \right\} \quad (19)$$

A variety of algorithms are available in literature to estimate the MLE in eq. (19). We refer users to [16] for a thorough discussion of estimation methods. For more details about the *copula* theory readers are referred to [17].

Predictions From A Copula: For a new observation \mathbf{x} we have to predict y . For this we form the conditional first by

$$P(y|\mathbf{x}) = \frac{P(\mathbf{x}, y)}{\int_y P(\mathbf{x}, y) dy}. \quad (20)$$

Our predicted \hat{y} maximizes this conditional probability

$$\hat{y} = \arg \max_{y \in Y} P(y|\mathbf{x}). \quad (21)$$

Note that the term in the denominator of eq.(20) remains constant, hence for the purposes of finding the optimum we can ignore its evaluation. We simply evaluate this conditional for the entire range of Y in discrete steps and pick the value of $y \in Y$ that maximizes the conditional.

4 Evaluation Setup

To evaluate and compare our different algorithms, we acquired a variety of wind data from the state of Massachusetts. We downloaded the data from ASOS (Automated Surface Observing System) airport database which is public and has wind data from fourteen airports in Massachusetts collected over the last ten to twenty years. It is frequently used by the wind industry. The airports' locations are shown in Figure 4. We then acquired data from an anemometer positioned on the roof top of Boston's Museum of Science where a wind vane is also installed. These anemometers are inexpensive and consequently noisy. The museum is located amongst buildings, a river and is close to a harbor as shown in Figure 5. This provides us with a site that is topographically challenging. At this location we have approximately 2 years worth of data collected at a frequency of 1 sample/second with 10 minute averages stored in a separate database.

To derive the wind resource assessment we train using data from the first year. This data is split into three datasets we call D_3 , D_6 and D_8 . The split D_3 has data for 3 months. The split D_6 has 3 additional months for a total of 6 and D_8 has yet 2 more months for a total of 8. We divide each dataset and the second year's dataset further into 12 directional bins of equal sizes starting at

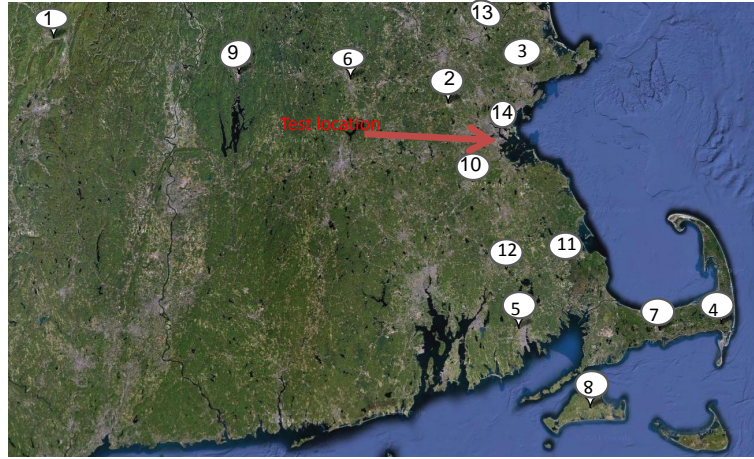


Fig. 4. Data is referenced from fourteen airport locations in the state of Massachusetts, USA.



Fig. 5. Red circles show location of anemometers on rooftop of Museum of Science, Boston.

Variable	Airport	Distance	(Lat, Long)
x_1	North Adams	151	(42.69°, -73.16°)
x_2	Bedford	18.7	(42.46°, -71.28°)
x_3	Beverly	20.6	(42.58°, -70.91°)
x_4	Chatham	88.5	(41.68°, -69.99°)
x_5	New Bedford	56.2	(41.67°, -70.95°)
x_6	Fitchburg	41.2	(42.55°, -71.75°)
x_7	Hyannis	72.7	(41.66°, -70.28°)
x_8	Lawrence	28.1	(42.71°, -71.11°)
x_9	Vineyard Haven	92.9	(41.39°, -70.61°)
x_{10}	Orange	84.7	(42.57°, -72.28°)
x_{11}	Norwood	23.2	(42.18°, -71.17°)
x_{12}	Plymouth	43.9	(41.96°, -70.68°)
x_{13}	Taunton	39.3	(41.87°, -71.01°)
x_{14}	Boston	4.9	(42.36°, -71.02°)

Table 1. The Boston Museum of Science is at position (N42.37°, W71.1°). Columns 1 and 2 show the airport names and their corresponding variable names in our multivariate model. Column 3 shows each airport’s line of sight distance in miles from the Boston Museum of Science. Column 4 shows the compass position of each airport.

compass point North (0°). We assume a linear regression model can be used to identify the wind direction at the Museum of Science. The second year’s dataset becomes our “ground truth” - the true wind resource assessment of the site and allows us to evaluate and compare the different techniques. We estimate a Weibull distribution model of it for this purpose. As a measure of predictive accuracy we compare the final estimated Weibull distribution to the ground truth distribution using Kullback-Leibler (KL) divergence. The lower this value, the more accurate the prediction:

$$D_{(Y||\hat{Y})} = KL(P_Y(y)||P_{\hat{Y}}(\hat{y})). \quad (22)$$

KL divergence derives the distance between two probability distributions:

$$D_{KL}(P_Y(y)||P_{\hat{Y}}(\hat{y})) = \sum_i P_Y(y = i) \ln \frac{P_Y(y = i)}{P_{\hat{Y}}(\hat{y} = i)} \quad (23)$$

For baseline comparison, we also developed a linear regression model which is used quite extensively in wind resource assessment [2, 8].

5 Results and discussion

In this section, we present the results of comparing the described wind resource assessment techniques on data acquired from the roof top anemometers at the Boston Museum of Science. We also examine the improvement in performance of each of the algorithms as more data is made available to each one in the forms of increasing training data from 3 to 6 to 8 months.

5.1 Comparison of algorithms

First we compare algorithms when the same amount of data is available to each one of them for modeling. Results are presented in Figures 6 to 8 for datasets D_3 , D_6 and D_8 respectively. Each plots shows the KL distance between the ground truth distribution and the distribution estimated based on the predictions provided by each technique for the year 2 dataset per bin. We plot the KL distance for all 12 bins.

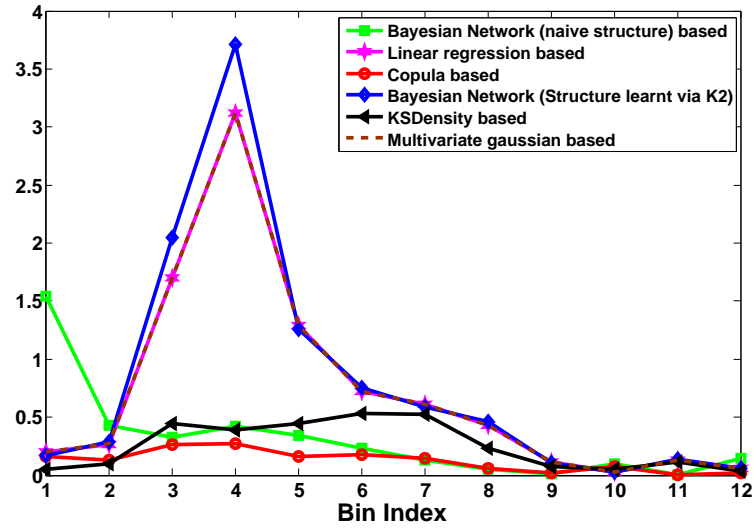


Fig. 6. Comparison of different techniques when 3 months worth of data is modeled and integrated with longer term historical data from 14 airports. These results were derived using D_3 and are compared with KL divergence distance to the Weibull distribution estimate of the second year of measurements at the Boston Museum of Science.

We notice that the copula modeling technique consistently performs better than the other four techniques. The graphical model technique which assumes a naive variable dependency structure performs second best though it demonstrates poor performance on the first bin. Its performance on this bin, however, improves as we increase the size of the dataset. One would expect the graphical model which has a learned variable dependency structure to outperform the one with naive structure assumptions. Here, except for the first bin, it does not. This may imply a better structure learning algorithm is necessary or that the one used needs further fine tuning. The latter possibility is likely because the structure learning algorithm K2 only looks at a fraction of all possible structures when it references an order of the variables. A more robust structure learning algorithm that does not assume order could potentially yield improvements.

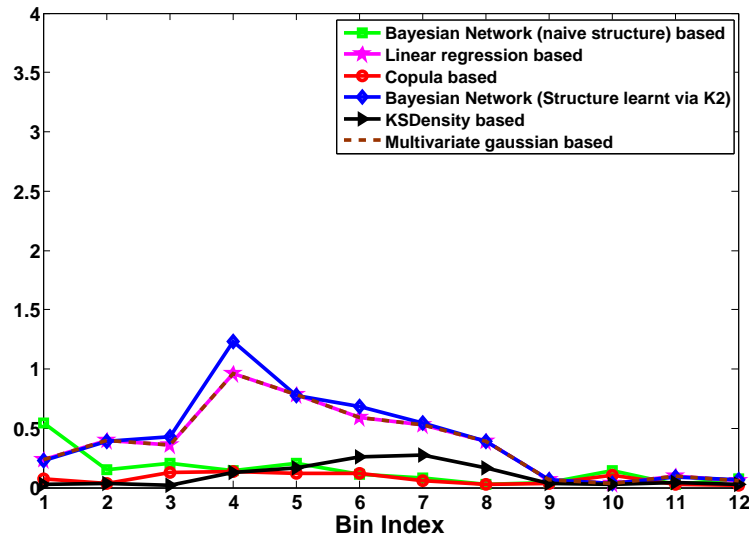


Fig. 7. Comparison of different techniques when 6 months worth of data is modeled and integrated with longer term historical data from 14 airports. These results were derived using D_6 and are compared with KL divergence distance to the Weibull distribution estimate of the second year of measurements at the Boston Museum of Science.

Linear regression is the worst performer of the all, but performs well when 8 months of data is available. This is consistent with many studies in the wind energy area where it has been found that for an accurate estimation of long term distribution 8 months worth of data is needed.

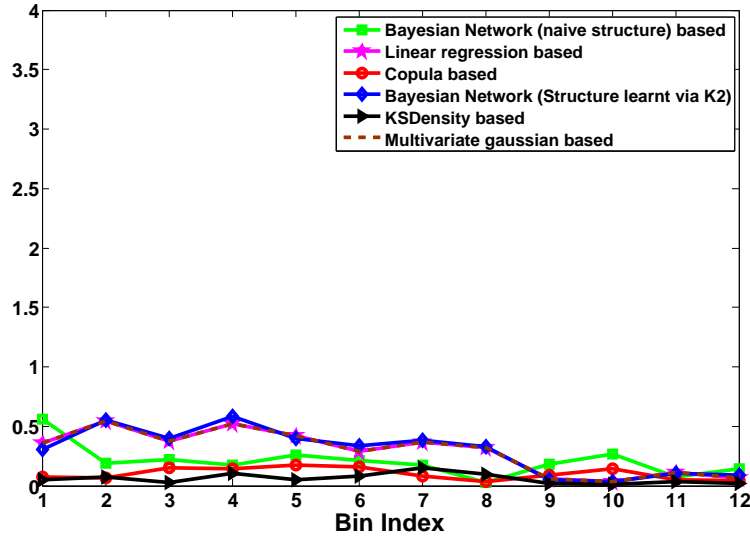


Fig. 8. Comparison of different techniques when 8 months worth of data is modeled and integrated with longer term historical data from 14 airports. These results were derived using D_8 and are compared with KL divergence distance to the Weibull distribution estimate of the second year of measurements at the Boston Museum of Science.

5.2 Increasing the data available for modeling

We now examine how each technique approaches robustness when less data is made available to it for modeling. Figure 9 plots each technique in isolation when it is modeled using 3, 6 or 8 months of data (datasets D_3 , D_6 and D_8) respectively.

We observe that not only was the copula modeling technique superior overall, its performance did not suffer greatly with decreasing amounts of data available for modeling. The graphical model with naive variable structure overcame its weak performance predicting the first bin as more data was made available to it. Both linear regression and the graphical model with learned variable structure improved significantly as more data was made available to them.

6 Conclusions and Future Work

In this chapter we have provided a set of techniques for building a statistical model for wind resource assessment. Our goal with these techniques was to

- estimate the wind speed density with as minimal site collected data as possible
- estimate as accurately as possible with minimal cost to support inexpensive site sensing.

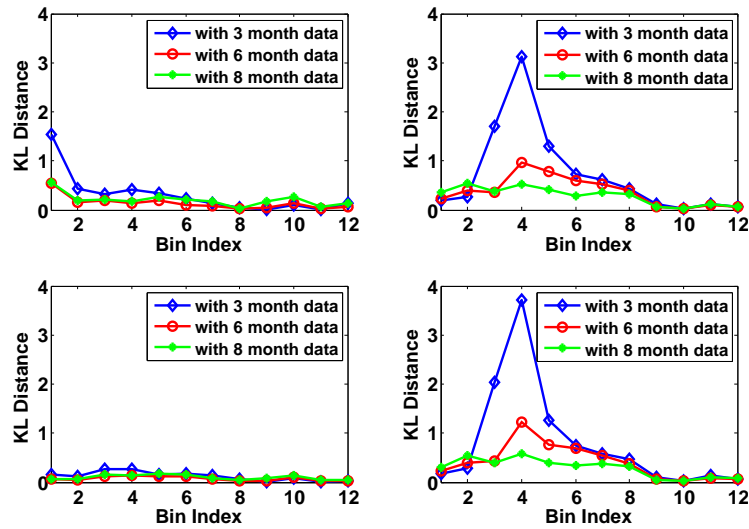


Fig. 9. Top(left): Estimation accuracy using graphical models with naive structure. Top (right): Estimation accuracy using linear regressions. Bottom (left): Estimation accuracy using *copulas*. Bottom(right): Estimation using Bayesian networks where structure is learned.

The ability to generate accurate estimates with as minimal data as possible and with as cheap sensing as possible is critical for wind resource assessment during the initial phases of wind farm planning. For community or urban wind energy projects anemometer sensing provides a cost effective way to estimate wind resource.

In definition the techniques are different in terms of whether they are parametric or not and whether they incorporate all variables into the joint distribution. This seems to have effect on the accuracy of the model in wind resource assessment. The copula modeling is more accurate than all other techniques.

We further analyzed the performance of the techniques when different amounts of data is made available to the modeling step. The technique based on Copula theory performs well even when only minimal data (3 months) is available. Another much simpler technique based on graphical models produces competent results as well.

Through this chapter we emphasize the need for exploration of a variety of statistical modeling techniques. A variety of additional advances can be made for each of the techniques presented in this chapter. These include but are not limited to:

Copula based functions: One can estimate the marginals using non-parametric kernel density functions to prepare input to a *copula*. One could also explore

systematically which *copula* is better however the extent to which tail behavior needs to be accurately modeled is open to debate. A variety of *copulas* are documented in statistics literature opening opportunities for further study.

Bayesian network functions: A more advanced structure learning algorithm could be used to estimate the Bayesian network structure. Different parameter estimation techniques could be explored.

Copula Bayesian networks : The network structure could be sought while forming the conditional at each node via a copula based multivariate density function. This concept has been recently explored for classification problems in machine learning, see [18]

Acknowledgements

We thank Steve Nichols (IIT Project Manager) and Marian Tomusiak, (Wind Turbine Lab Analyst), of the Museum of Science, Boston for assisting us in data acquisition and assessment. We thank the MIT Energy Initiative for its sponsorship of Adrian Orozco who prepared and synchronized the data collected from the Museum of Science as an Undergraduate Research Assistant.

References

1. Gross, R.C., Phelan, P.: Feasibility study for wind turbine installations at museum of science, boston. Technical report, Boreal Renewable Energy Development (October, 2006)
2. Bass, J., Rebbeck, M., Landberg, L., Cabré, M., Hunter, A.: An improved measure-correlate-predict algorithm for the prediction of the long term wind climate in regions of complex environment. (2000)
3. Bailey, B., McDonald, S., Bernadett, D., Markus, M., Elsholz, K.: Wind resource assessment handbook: Fundamentals for conducting a successful monitoring program. Technical report, National Renewable Energy Lab., Golden, CO (US); AWS Scientific, Inc., Albany, NY (US) (1997)
4. Lackner, M., Rogers, A., Manwell, J.: The round robin site assessment method: A new approach to wind energy site assessment. *Renewable Energy* **33**(9) (2008) 2019–2026
5. Encraft: Warwick wind trials final report. Technical report, Encraft LLC, Holly Walk, UK (2009)
6. Shaw, S.: Progress report on small wind energy development projects receiving funds from the Massachusetts Technology Collaborative (MTC). Cadmus Group Inc, Waltham, MA (2008)
7. Wagner, M., Veeramachaneni, K., Neumann, F., O’Reilly, U.: Optimizing the layout of 1000 wind turbines. In: Scientific Proceedings of European Wind Energy Association Conference (EWEA 2011). (2011)
8. Rogers, A., Rogers, J., Manwell, J.: Comparison of the performance of four measure-correlate-predict algorithms. *Journal of wind engineering and industrial aerodynamics* **93**(3) (2005) 243–264
9. Chan, C., Stalker, J., Edelman, A., Connors, S.: Leveraging high performance computation for statistical wind prediction. In Proceedings of WINDPOWER 2010 (2010)

10. Frank, E., Trigg, L., Holmes, G., Witten, I.: Technical note: Naive bayes for regression. *Machine Learning* **41**(1) (2000) 5–25
11. Koller, D., Friedman, N.: Probabilistic graphical models: principles and techniques. The MIT Press (2009)
12. Heckerman, D., Geiger, D., Chickering, D.: Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning* **20**(3) (1995) 197–243
13. Cooper, G.F., Herskovits, E.: A bayesian method for the induction of probabilistic networks from data. *Mach. Learn.* **9**(4) (1992) 309–347
14. Murphy, K., et al.: The bayes net toolbox for matlab. *Computing science and statistics* **33**(2) (2001) 1024–1034
15. Burton, T., Sharpe, D., Jenkins, N., Bossanyi, E.: Wind energy: handbook. Wiley Online Library (2001)
16. Iyengar, S.: Decision-making with heterogeneous sensors-a copula based approach. PhD Dissertation (2011)
17. Nelsen, R.: An introduction to copulas. Springer Verlag (2006)
18. Elidan, G.: Copula bayesian networks. (2010) 559–567
19. Eaton, M., Euclid, P., Library, C.U., Press, D.U.: Multivariate statistics: a vector space approach. Wiley New York (1983)

Appendix

Below we describe how to derive the conditional density function parameters for y given \mathbf{x}_k under the assumption that the joint is modeled as a normal. We first partition the mean and the covariance matrix for the joint distribution of \mathbf{z} as follows:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{bmatrix} \quad (24)$$

with sizes

$$\begin{bmatrix} 1 \times 1 \\ m \times 1 \end{bmatrix} \quad (25)$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{bmatrix} \quad (26)$$

with sizes

$$\begin{bmatrix} 1 \times 1 & 1 \times m \\ m \times 1 & m \times m \end{bmatrix} \quad (27)$$

then, the distribution of y conditional on $\mathbf{x} = \mathbf{x}_k$ is multivariate normal with [19]

$$\overline{\boldsymbol{\mu}_{y|\mathbf{x}_k}} = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} (\mathbf{x}_k - \boldsymbol{\mu}_x) \quad (28)$$

and covariance matrix

$$\overline{\boldsymbol{\sigma}_{y|\mathbf{x}_k}} = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{xy}. \quad (29)$$