

ABSTRACT

Renewable energies are of paramount importance when it comes to energy consumption. They provide reliable power sources and fuel diversification, which improves energy security and helps reduce the risk of fuel spillage and the need for imported fuels. They also help preserve the country's natural resources. Wind energy production must be increased to be able to use them at high capacity. Wind speed is one of the most reliable sources of clean and sustainable electricity supply. The use of wind speed is one of the crucial factors that significantly contribute to the renewal of energy, and it provides an enormous benefit by increasing wind power generation. But the major problem with this increase is the uncertainty inherent in the wind speed. For too long, the issue of uncertainty has been at the center of deep thinking. Scientists, researchers, and academics are working on the central question of how wind speed should be used while considering its nature of uncertainty. This uncertainty is due to the day/night cycle caused by the earth's rotation and the seasonal changes due to the tilt of the earth's axis, both of which cause changes in wind speed. So far, the wind speed uncertainty has been presented as a probability distribution. However, the uncertainty of these wind speed models has not yet been considered. This dissertation uses three approaches to thoroughly analyze fifteen wind speed data variables collected from the National Renewable Energy Laboratory. The data are split into three different sample sizes, namely (hourly 8760, daily 365, and weekly 53). Firstly, the dissertation applies the non-Bayesian and the Bayesian approaches to study linearity among the data. In this first approach, a correlation matrix method is implemented to select the most correlated variables and use the highest correlated variable among them as the dependent variable. After selecting the dependent variable from the correlation matrix method, we proceeded by applying a Random Forest machine learning technique for the feature selections and considered the most important features to be used as independent variables in both the non-Bayesian and Bayesian regression models. Secondly, to ensure that the nature of the uncertainty is carefully analyzed and minimized, we analyzed the variables again to determine their dependencies. We have applied different vine models such as R-vine, C-vine, and D-vine copulas to analyze the variables' dependencies on each data size (i.e., hourly, daily, and weekly). The empirical pairwise Kendall Tau values and pairwise copula families are used to assessing the data's dependencies after the data are normalized. Loglikelihood, AIC, and BIC are used as measurement tools to select the best-fitted model. Finally, the dissertation used the Bayesian Moving Average method to predict the wind data analyzed through the linearity and dependencies approaches. Only the medium dataset is used due to promising results when obtained from the precedents analysis. We ran three different equations. We first determined the most important variables with a higher coefficient based on Post Mean. Second, we used different priors' models and ran an MCMC model to ensure the best fit of the models. Thus, we used the posterior coefficient density to analyze the entire posterior distribution of the coefficients and compare the expected values of the coefficients. For the prediction, we used the last 165 days of the medium dataset (daily 365).

RESEARCH OBJECTIVES & METHODS

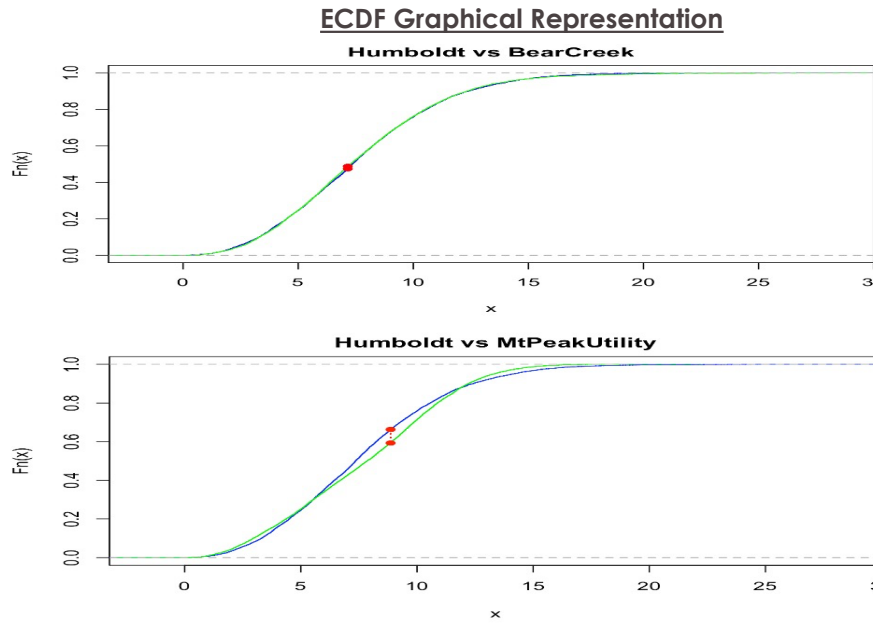
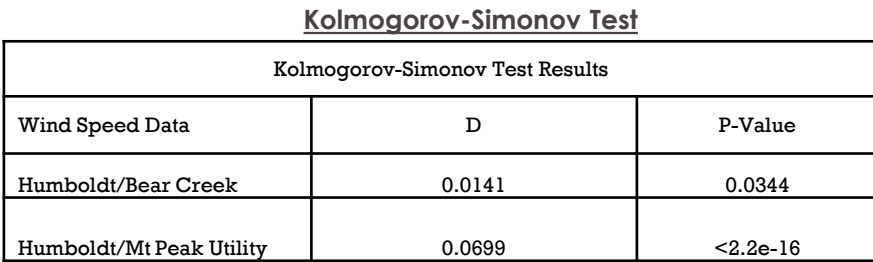
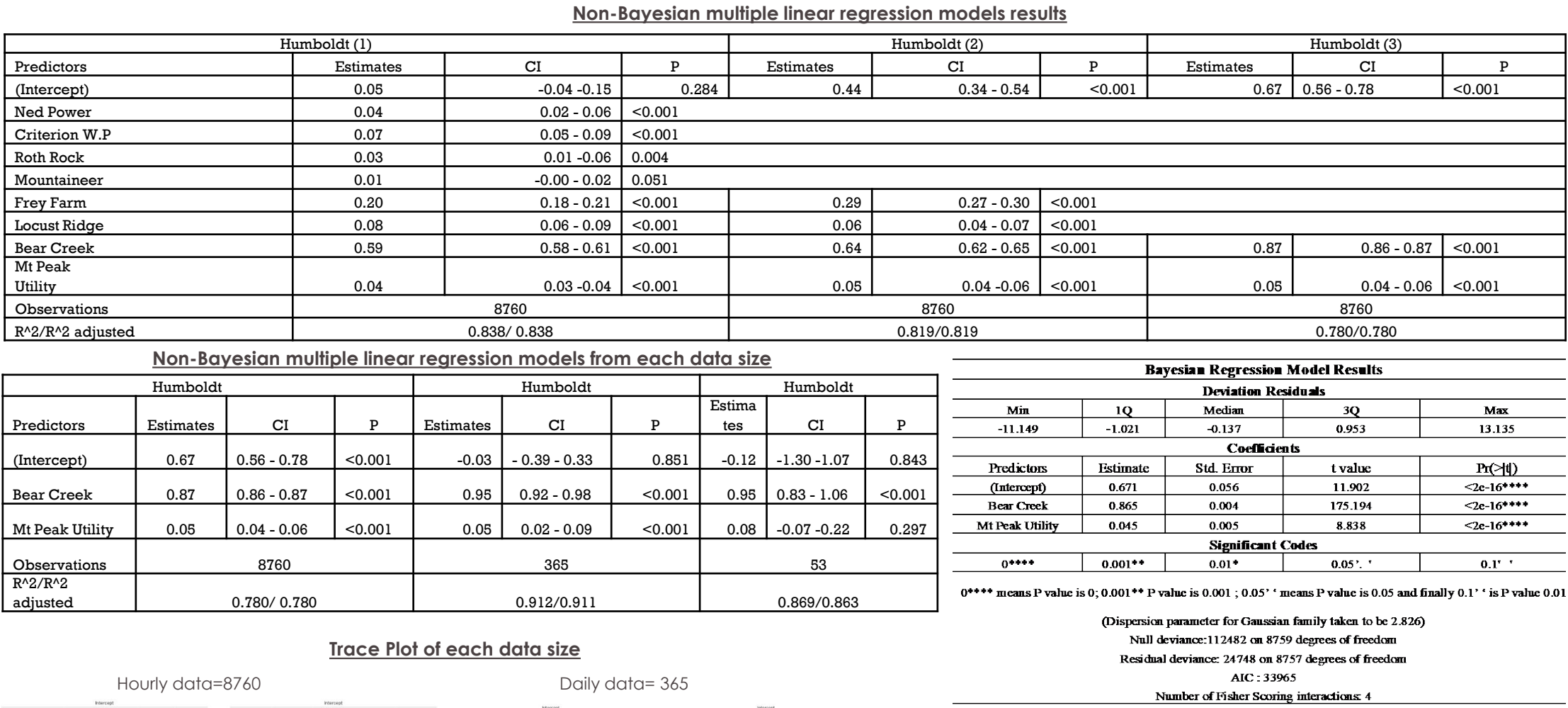
This dissertation analyses wind speed data variables, including their nature of uncertainty, using three statistical methods. The main objective is to thoroughly investigate the data and ensure that uncertainty is tackled when predicting the data. The three statistical methods are defined as follows: The first is the linear regression method based on a Bayesian and non-Bayesian model. Furthermore, this dissertation also focuses on the second approach, which is applying vine copulas to analyze dependencies among wind speed data. Finally, the prediction will be made from the Bayesian Model Average application. These three proposed models are integrated and analyzed to better predict the dependent variable while considering the nature of uncertainty. The following steps are consolidated:

**Step1: Non-Bayesian** [Correlation Matrix; Random Forest for features selections, Multiple Linear Regression; Imposition of Variance Inflation Factor(VIF) for model selection; Fitted Model selection; Kolmogorov Smirnov Test(KS); Prediction]. **Bayesian Model** [Fitted model selected; Priors and Maximum Likelihood Estimator; Markov Chain Monte Carlo (MCMC)].

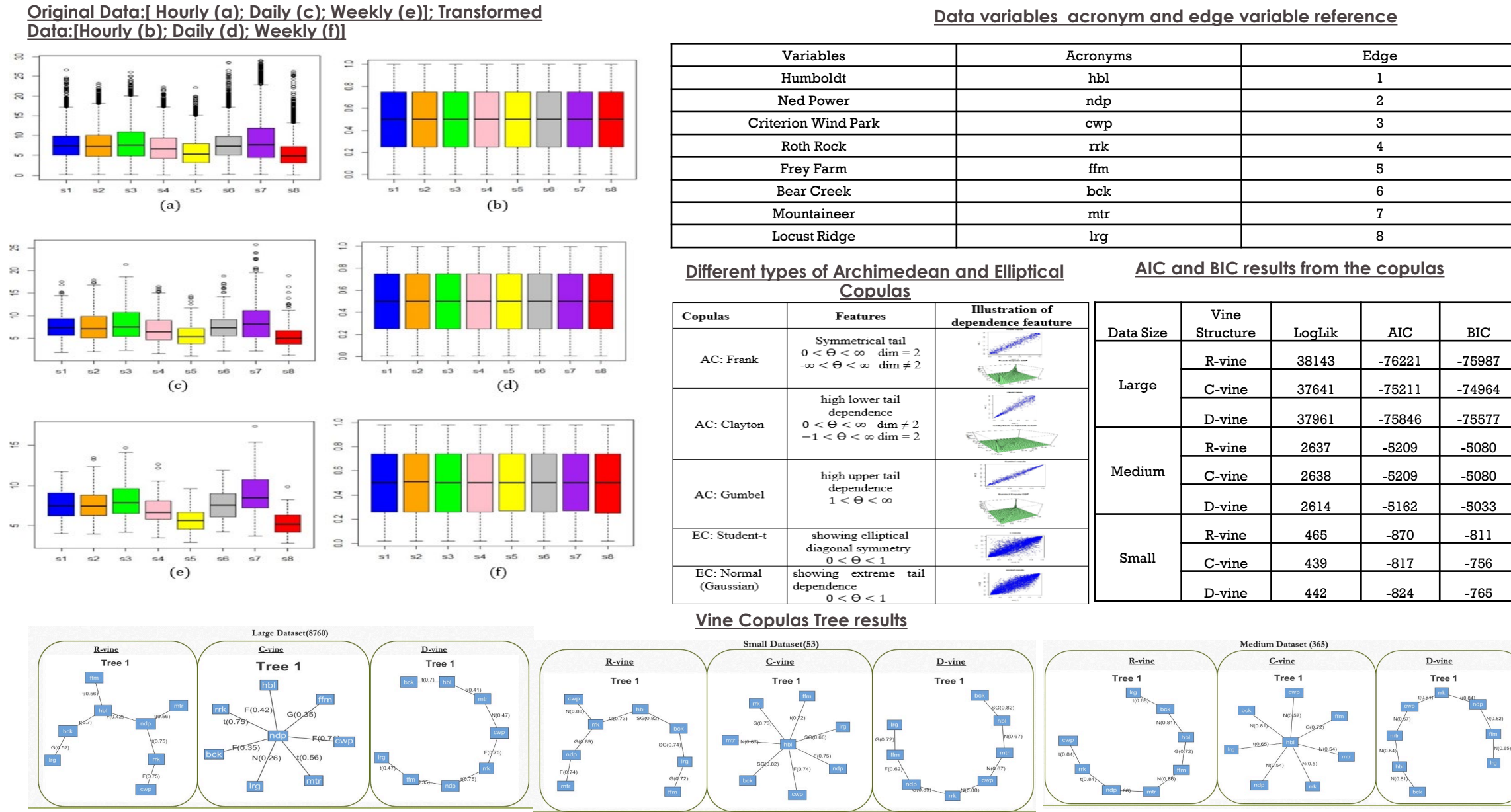
**Step2: Vine Copulas** [Transform data to copulas data; Kendall's Tau applications; Archimedean and Elliptical copulas, AIC, BIC and Log Likelihood (Loglik) applications for vine selections; Evaluation of Vine Copulas family (R-vine, C-vine, and D-vine)].

**Step3: Bayesian Moving Average (BMA)**[ Applications of BMA on Daily (365); Canonical regression applications; Posterior Inclusion Probabilities(PIP) and Post Mean (PM) for coefficients selections and Model comparisons; Forecast based on the predictive density model using 165 days.]

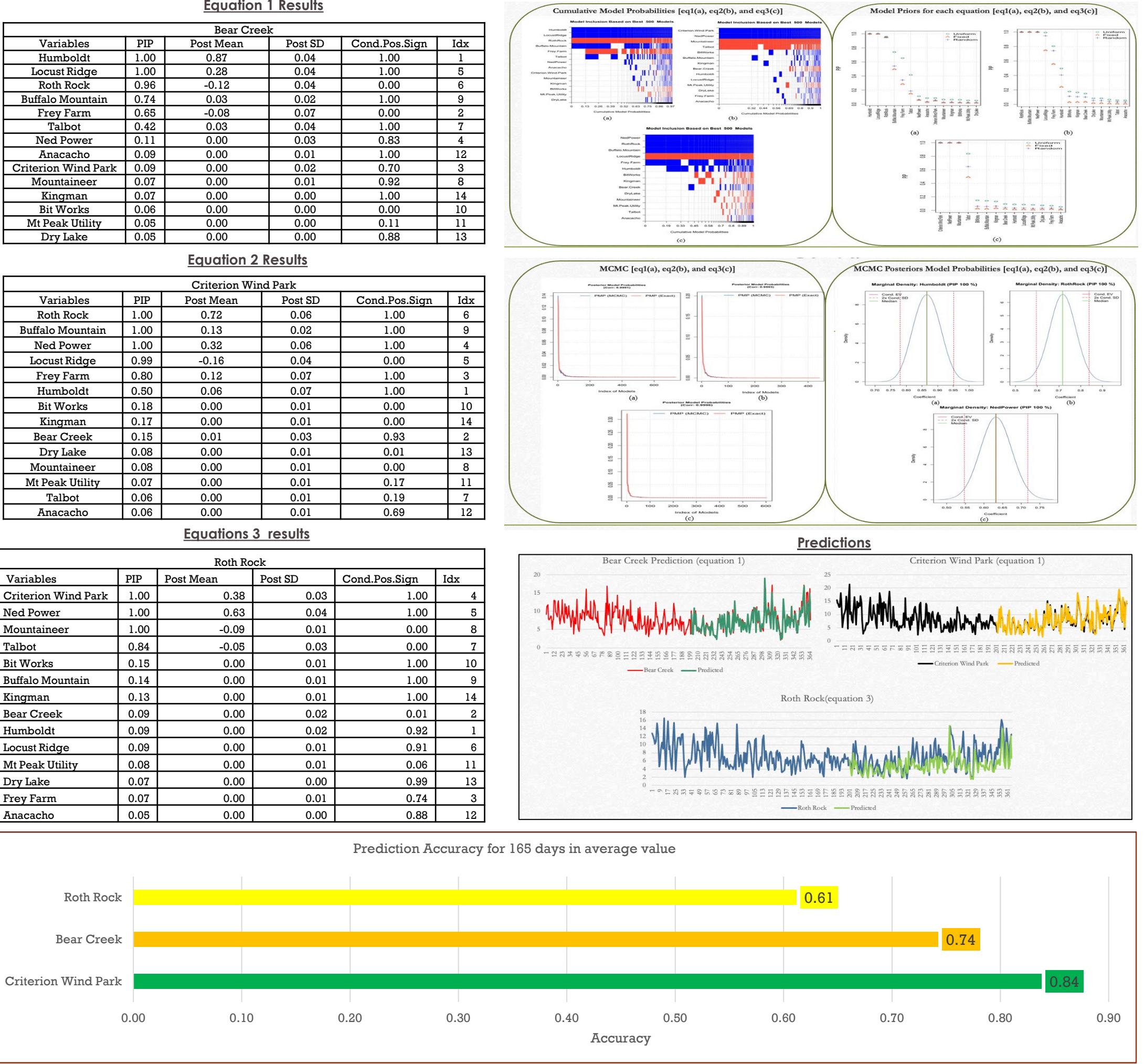
BAYESIAN AND NON-BAYESIAN RESULTS



VINE COPULAS RESULTS



BAYESIAN MOVING AVERAGE RESULTS



CONCLUSIONS

To analyze the nature of uncertainty in wind speed data, we have proposed three simultaneous statistical models which are (i) Bayesian and non-Bayesian models for the linear model, (ii) vine copula models for dependence analysis, and (iii) Bayesian moving average for implementing the prediction. We applied the non-Bayesian and the Bayesian approaches to analyze the multiple linear regressions for wind speed data and evaluate the differences between the two statistical methods. In contrast to several studies based on the Bayesian development models that have shown better results than the non-Bayesian methods, our study has shown that both the non-Bayesian and the Bayesian approaches are very much alike in the coefficients/parameters estimations. On the vine copulas approach, we used models such as R-vine copulas, and C and D-vine copulas. In the results, we have observed the dominance of the Ned Power variable over the other variables when it comes to the large dataset and Humboldt for both the medium and small datasets. The results have also shown that at the level of the likelihood estimation, there is almost a similarity between R-vine and C-Vine models in the medium dataset. We finally applied the BMA model using the medium dataset for prediction. We found that in the last 165 days, the real values and the predicted ones were very close and that the margin of error was insignificant.in some cases especially for Criterion Wind Park and Beak Creek. In the case of Roth Rock, our prediction was somewhat off the mark. The prediction accuracy of the Criterion Wind Park variable was 0.84, that of Bear Creek 0.74, and finally that of Roth Rock 0.61.

ACKNOWLEDGEMENTS

Thank you to the Department of Computer Science and Engineering and all the committee members.