ECAI 2010
H. Coelho et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-606-5-661

Adaptive Gaussian Process for Short-Term Wind Speed Forecasting

Xiaoqian Jiang* and **Bing Dong+** and **Le Xie#** and **Latanya Sweeney◇1**

Abstract. We study the problem of short term wind speed prediction, which is a critical factor for effective wind power generation. This is a challenging task due to the complex and stochastic behavior of the wind environment. Observing various periods in the wind speed time series present different patterns, we suggest a nonlinear adaptive framework to model various hidden dynamic processes. The model is essentially data driven, which leverages non-parametric Heteroscdastic Gaussian Process to model relevant patterns for short term prediction. We evaluate our model on two different real world wind speed datasets from National Data Buoy Center. We compare our results to state-of-arts algorithms to show improvement in terms of both Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

1 Introduction

Wind energy has been developed significantly throughout the world recently in order to achieve a low-emission electricity sector. However, many important issues emerge in the integration of wind farms in the power networks for the unity of commitment and control of power plants in electric power systems. The objective of electric energy system operation is to maintain the balance of total supply and demand under certain transmission constraints. Conventionally before the integration of large-scale intermittent resources such as wind, the operating philosophy has been controlling generation resources in order to fully balance the time-varying loads. The unpredictable part of loads is balanced in automated manner commonly referred to as automatic generation control (AGC). The introduction of large-scale wind energy into the existing power grid poses significant challenges to the amount of reserves and the amount of control efforts needed for balancing the system [14]. Therefore, an accurate wind speed prediction is essential for running the future power grids in a cost-effective and reliable manner.

The electric energy system operation is classified into several different time scales, namely, day-ahead dispatch (unit commitment), short-term dispatch, and real-time automatic generation control. Day-ahead dispatch usually occurs 24 hour ahead of the operating time, and determine the hourly ON and OFF status of all the generating units in the system. When it comes to a short-time, economic dispatch program takes place every 30 minutes, which minimizes the total generation cost within the transmission constraints. Daily and hourly wind forecasts are useful information in the decision process of Day-ahead Dispatch, whereas short-term 30 minute-ahead wind forecasting information is critical for efficient economic dispatch. In this paper we focus on the <30-minute ahead wind prediction.

2 Literature Review

Various time series regression approaches have been applied to the problem of short term wind speed prediction. Sfetsos applied various forecasting techniques to predict mean hourly wind speed . These techniques include linear models (ARMA), feed forward and recurrent neural networks (NNT), adaptive neuro-fuzzy inference systems and neural logic networks [12]. The results show that NNT methods are better than other linear and non-linear models within 5% in terms of root mean square error (RMSE). Hunt and Nason treated wind speed data as nonstationary time series and applied wavelets method to predict 10-minute ahead wind speed data over 21 days from four different locations in U.K. [6] . They also compared with linear baselines and showed their wavelet model is 20% better. Palomares-Salas et al. compared two models, ARMA and NNT, for short term wind speed forecasting (10 minutes, 1 hour, 2 hours and 4 hours) and the results are very similar with average RMSE from 0.57 to 1.55 [9]. However, applying these approaches without the context of stochastic wind nature makes the model very sensitive to the controlling parameters and often do not generalize well to novel observations.

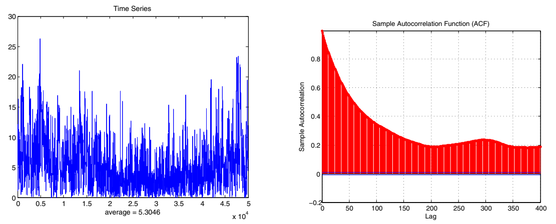
Alternatively, as the wind speed has momentum over time, it can be treated as transitional behavior and modeled as a Markov Process. Tore et al. (2001) used first order Markov chain models for synthetic generation of hourly wind speed time series in the Turkey [13]. [15] modeled both hourly wind speed and wind direction data based on Markov chains. Shamshad et al. have generated hourly wind speed data using first and second order Markov chains and compared the first and second order Markov chains using wind speed data measured from two different regions in Malaysia [11]. In their study, it was concluded that the wind speed behavior slightly improves by increasing the Markov model order. Most recently, Hocaoglu et al. applied Mycielski algorithm to predict the hourly wind speed data for three locations in Turkey in the sense of forecasting future values of wind data by analyzing the repeated data in the history of the whole data [4].

This study presents a different and novel model to tackle the short term wind speed prediction problem. Our approach is essentially data driven, which leverages non-parametric algorithms to locate and model relevant sub-sequences of observation adaptively. Although the idea of the proposed model also depends on learning from past samples, unlike the Markov approach of building transition probabilities, it considers the past data samples together during the prediction.

1 *◇School of Computer Science, Carnegie Mellon University, +School of Architecture, Carnegie Mellon University, # ECE department, Carnegie Mellon University, contact email: xiaoqian@cs.cmu.edu

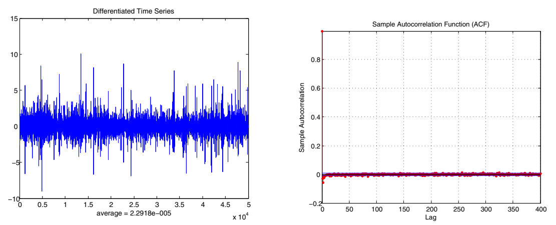
3 Data Analysis

We record two wind speed datasets from the National Data Buoy Center [1], the New York area wind speed dataset (NY-ISO) and the Calumet, IL area wind speed dataset (Calumet-ISO). The NY-ISO dataset contains $52,262$ measurements corresponding to wind speed in meters per second (m/s), acquired at each 10 minutes, covering an entire year from 2006 to 2007. The Calumet-ISO dataset contains $10,471$ measurements, acquired every 6 minutes, ranging from 1/1/2010 to 2/16/2010.



(1) (2)

(a) Wind speed time series and autocorrelation. (1) Wind speed time series, (2) autocorrelation



(1) (2)

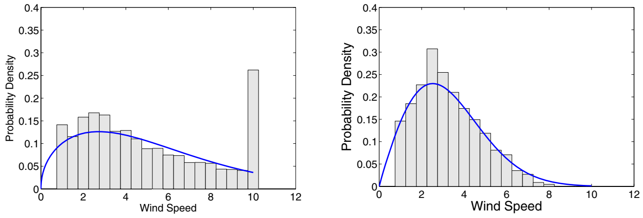
(b) Differentiated wind speed time series and autocorrelation. (1) Differentiated wind speed time series, (2) Autocorrelation of differentiated time series

Figure 1: Autocorrelation analysis for time series wind speed data.

The time series behavior and the autocorrelation of NY-ISO are depicted in Figure 1a. The autocorrelation coefficient starts at a high value and decays quickly as the time increases. After 50 minutes, the decay slows down and becomes stable. The differentiated time series and corresponding autocorrelation are depicted in Figure 1b. The near-zero autocorrelation of the differentiated time series indicates that the short term autocorrelation is non-linear.

Next, we fit the both datasets with a two-parameter Weibull distribution, as illustrated in Figure 2. The Anderson-Darling test shows that neither wind speed datasets satisfies Weibull distribution as opposed to some previous literature claimed for their data [2, 15]. Thus, moving average as the Maximum Likelihood Estimator of a single distribution is not suitable for wind speed prediction. The statistical analysis also reveals the following facts:

1. Future wind speed is strongly correlated to the very short term historical data.
2. An appropriate regression function should be non-linear.
3. Long term historical observation contains useful information for forecasting.



(a) (b)

Figure 2: Fit Wind Speed time series data with a single Weibull distribution. (a) Fitted PDF (blue line: 2-parameter Weibull) of the NY-ISO dataset (b) Fitted PDF of the IL-ISO dataset. Both null hypotheses are declined at the significance level 0.05 use Anderson-Darling test.

4 Methodology

4.1 Notations

Motivated by the findings, we propose a novel model, Adaptive Gaussian Process (AGP), to model wind speed. Let's start with the notations in the following table.

Variables	Summary
$\mathbf{X} = [X_t]\|_{t=1}^{T-1}$.	The raw time series data
$\mathbf{P} = \{P_{t'-1}, X_{t'} \| P_{t'} \in \mathcal{R}^w, \|\|P_{t'} - P_t\|\| \in \Omega_K\}$.	The *locality* of P_t, k most similar patterns to P_t
$\mathbf{h}$=$\{h_1, \ldots, h_t, \ldots, h_T\}$	Random variables correspond to the P_t
w	The sliding window
$P_{t-1} = [X_{t-w} : X_{t-1}] \sim X_t$	The local pattern associated with X_t
Ω_K	The space of *locality*
$h_t = h(P_t)$	Gaussian Process function evaluated at t-th local pattern
$m(h_t)$	Mean function
$k(h_t, h_{t'})$	Kernel function
α, Λ	Model parameters
$K = K(\mathbf{P}, \mathbf{P})$	Kernel matrix
$R = diag(\mathbf{r}), \mathbf{r} = (r(\mathbf{P}))^T$	Heteroscdastic per-sample noise

Table 1: Noations of the variables.

4.2 Adaptive Gaussian Process

We believe a single regression model cannot capture the dynamic wind speed patterns under various environments, refer to Figure 4. In fact, models constructed with irrelevant data might harm the performance. For example, including wind speed of the night may decrease the day-time prediction accuracy because the wind speed is usually stronger at night. Though data is collected sequentially, models should be constructed locally. Hence, we propose a dynamic framework that constructs models "locally" with strongly relevant patterns.

Let wind speed at time t be X_t and the entire time series can be represented as a collection of X_t, $\mathbf{X} = [X_t]|_{t=1}^{T-1}$. We define a sliding window as w and aggregate $P_t = [X_{t-w} : X_{t-1}]$ to construct a *local pattern* $P_t \in \mathcal{R}^w$ for the target X_t and denote this relation as $P_t \sim X_t$. We define the *locality* of each wind speed observation X_t to be its K most similar local patterns

$$\mathbf{P} = \{P_{t'}, X_{t'} | P_{t'} \in \mathcal{R}^w, ||P_{t'} - P_t|| \in \Omega_K\}.$$

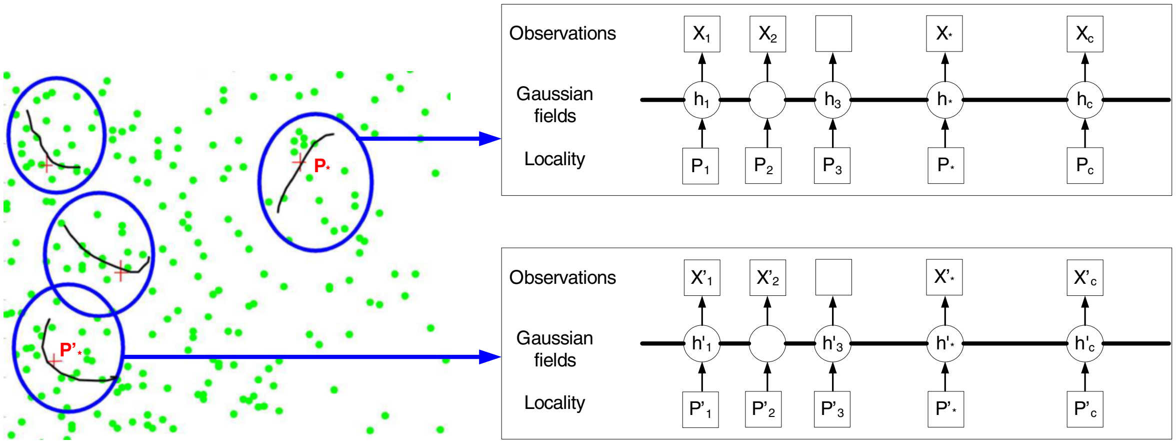


Figure 3: AGP illustration: it trains separate Gaussian Processes on the K nearest neighbors of each target P_*, denoted as *locality* $\mathbf{P}$ (preserving the original distance metric), and directly obtains regression boundary. Note the $+$ sign corresponds to a local pattern and the blue circle indicates its *locality*. On the right column is the graphical models for GP regression. X, P represent observed variables and h represent unknowns. The thick horizontal bar means the hidden variables are fully connected. Due to the marginalization property, an observation P_t is conditionally independent of all other nodes given the corresponding latent variable, h_t.

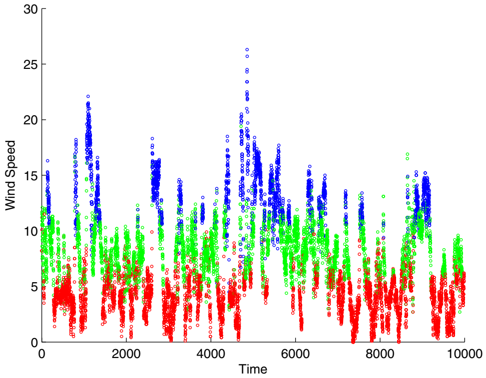


Figure 4: Different periods in the wind speed time series present diverse patterns, thus various dynamical processes should be considered. Note the blue, green and red parts correspond to different patterns in wind speed time series. For illustration purpose, we show 3 major *localities* but our model generates significantly more of them.

For each observation X_t, we construct a *locality* for time $t+1$ and use L_t to forecast X_{t+1}. We start from a coarse and quick categorization, followed by successive finer but slower regression. Our proposed approach uses nearest neighbor at an initial pruning stage and leverages powerful regression techniques on the smaller but more relevant set of examples that require careful attention, refer to Figure 3 for details. The regression model is a variation of Gaussian Process (GP) [10] with relaxed noise assumption. Basically, a GP represent current wind speed X_t by the its *local pattern* through a regression model with the homoscedastic Gaussian noise,

$$X_t = h(P_t) + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma^2), \ 2 \leq t \leq T, \tag{1}$$

where $h(\cdot)$ is a nonlinear function such that $h(P) = \phi(P)^T w$. Note $\phi(\cdot)$ projects the inputs into some high dimensional space using a set of basis feature space functions and a linear model is applied in this space instead of directly on the inputs themselves. In this paper, we use RBF functions for $\phi(\cdot)$ and let $k(P, P') = \phi(P)\phi(P')$.

$$k(P, P') = \alpha^2 \exp\left(-\frac{1}{2}(P - P')\Lambda^{-1}(P - P')\right), \tag{2}$$

we call $k(\cdot, \cdot)$ a kernel function where α and Λ are model parameters. The kernel function is almost α^2 between variables when the corresponding inputs are close, and decreases quickly as their distance in the input space decreases. Such specification of kernel function implies the prior on a distribution of functions. For instance, we draw samples from a distribution of function evaluated at any number of points, e.g. $\mathbf{P}$, and write out the corresponding function distribution,

$$h(\mathbf{P}) \sim N(0, k(\mathbf{P}, \mathbf{P})). \tag{3}$$

A Gaussian process thus defines a Gaussian distribution on the function space $h(\cdot)$, assuming it is fully determined by its mean function and covariance function

$$\begin{aligned} m(P) &= E[h(P)] = 0, & (4)\\ k(P, P') &= E[(h(P) - m(P))(h(P') - m(P'))]. & (5) \end{aligned}$$

An important marginalization property of the Gaussian Process is that when the GP specifies $(P_1, P_2) \sim N(\mu, \Sigma)$, it must also specify $P_1 \sim N(\mu_1, \Sigma_{11})$, where Σ_{11} is the relevant sub-matrix of Σ. Thus, examination of a larger set of variables does not change the distribution of a smaller set. This consistency requirement is automatically fulfilled if the covariance function specifies entries of the covariance matrix [8]. Hence, it is easy to show that when $(h(\mathbf{P}), h(P_T)) \sim N(m([\mathbf{P}, P_T]), k([\mathbf{P}, P_T], [\mathbf{P}, P_T]))$, then the posterior predictive distribution of the function value $h_T = h(P_T)$ for a future wind speed *local pattern* P_T is still Gaussian,

$$h_T \sim N(m(P_T), k(P_T, \mathbf{P})) \tag{6}$$

with the following mean and the variance,

$$\begin{aligned} m(P_T) &= E[h_T] = k_T^{'}(K + \sigma^2 I)^{-1}\mathbf{X} & (7)\\ k(P_T, \mathbf{P}) &= var[h_T] = k_{TT} - \mathbf{k}_T^{'}(K + \sigma^2 I)^{-1}\mathbf{k}_T, & (8) \end{aligned}$$

where $k_{TT} = k(P_T, P_T)$, $\mathbf{k}_T = k(\mathbf{P}, P_T)$ and $K = K(\mathbf{P}, \mathbf{P})$ is the kernel matrix for the training data. Remember GP assumes

$i.i.d$ noise σ^2 for each *local pattern* associated with *locality* $\mathbf{P}$. In practice, this assumption suffers poor performance as it overlooks the complex stochastic behavior of the wind speed environments. To capture the high variability in wind, we must assume per sample noise. To this end, we model the noise by a function of P, thus we get a Heteroscdastic regression problem, where the noise rate is not assumed constant. By placing a Gaussian process prior on f and assuming a noise rate function $r(P)$, the predictive distribution $P(h_T|P_T, \mathbf{P}, \mathbf{X})$ at the query points P_T is a multivariate Gaussian distribution,

$$h_T|P_T, \mathbf{P}, \mathbf{X} \sim N(\bar{h}_T, cov(h_T)) \tag{9}$$

$$\bar{h}_T = k_T^{'}[K+R]^{-1}\mathbf{X}, \tag{10}$$

$$cov(h_T) = k_{TT} - r_T - \mathbf{k}_T^{'}[K+R]\mathbf{k}_T, \tag{11}$$

where $R = diag(\mathbf{r})$ with $\mathbf{r} = (r(\mathbf{P}))^T$ and $r_T = r(P_T)$. We use an independent GP to model the noise levels, so that this r-process is governed by a different covariance function k_r, parametrized by θ_r. Since the noise rates r_t are independent latent variables in the combined regression model, the predictive distribution for h_T, the vector of regressands at points P_T changed to,

$$P(h_T|P_T, \mathbf{P}, \mathbf{X}) = \int\int P(h_T|P_T, \mathbf{P}, \mathbf{X}, \mathbf{r}, r_T) P(\mathbf{r} r_T|P_T, \mathbf{P}, \mathbf{X}) d\mathbf{r} dr_T. \tag{12}$$

Given $(\mathbf{r}, r_T)$, the predictive process is Gaussian with specified mean and variance. The real difficult is thus $P(\mathbf{r}, r_T|P_T, \mathbf{P}, \mathbf{X})$ because it makes the integral difficult to handle analytically. [3] and [7]suggested Monte Carlo approximation and full Bayesian treatment, respectively. But both computation is quite time consuming. A recent paper [5] suggested an iterative procedure to estimate the most likely per sample noise and showed good empirical performance.

We suggest an easy way to estimate with the two rounds of optimizations. First, we assume a homogeneous Gaussian noise ϵ with variance σ^2 to learn an ordinary Gaussian Process $\hat{h}$ with kernel K_o. Then, we estimate the covariance on each *local pattern* from $\dot{\mathbf{P}} = \mathbf{P} \cup P_T$. Note Eq. 11 does not depend on the P_T and we can thus estimate the covariance error on both *locality* $\mathbf{P}$ and the P_T. Next, we plugin the covariance error of $\dot{\mathbf{P}}$ to form the new kernel.

$$K_n = K_o - \sigma^2 I_N + \sigma_{\hat{h}}^2(\dot{\mathbf{P}}) I_N. \tag{13}$$

Finally the value of the future wind speed X_{T+1} is estimated as the mean of the posterior

$$\tilde{X}_{T+1} = k_T^{'}(\hat{K}_n + \sigma^2 I)^{-1}\mathbf{X} \tag{14}$$

where $\hat{K}_n$ is composed of the rows and columns of K_n corresponding to $\mathbf{P}$. We can add an small noise variance σ^2 to our new function h_n to give robustness to the inversion of $\hat{K}_n$.

5 Experiments and Discussion

Direct comparison with other literature is difficult due to the lack of common test dataset and test-training partitions. We thus implement three recently proposed approaches in the literature and compared their performance with ours with the two public datasets, NY-ISO and Calumet-ISO.

- Method 1: Autoregressive moving average (ARMA) model [9] .
- Method 2: Second order Markov Chain [11].
- Method 3: Mycielski algorithm [4].

The model efficacy is compared in Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{15}$$

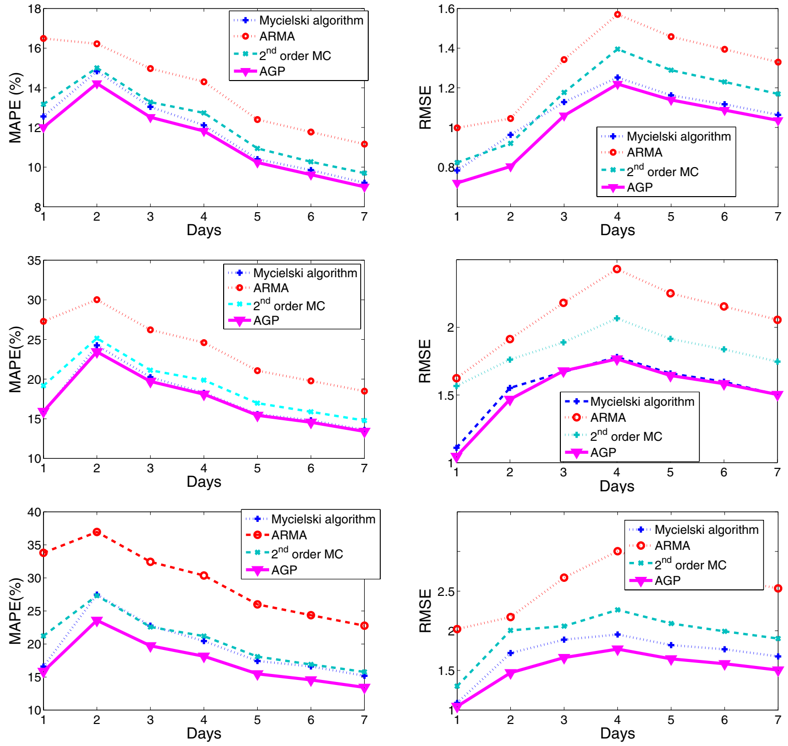
$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2}, \tag{16}$$

where A_t is the actual value and F_t is the forecast value. We conduct the experiments on a 2Ghz desktop with Matlab environment.

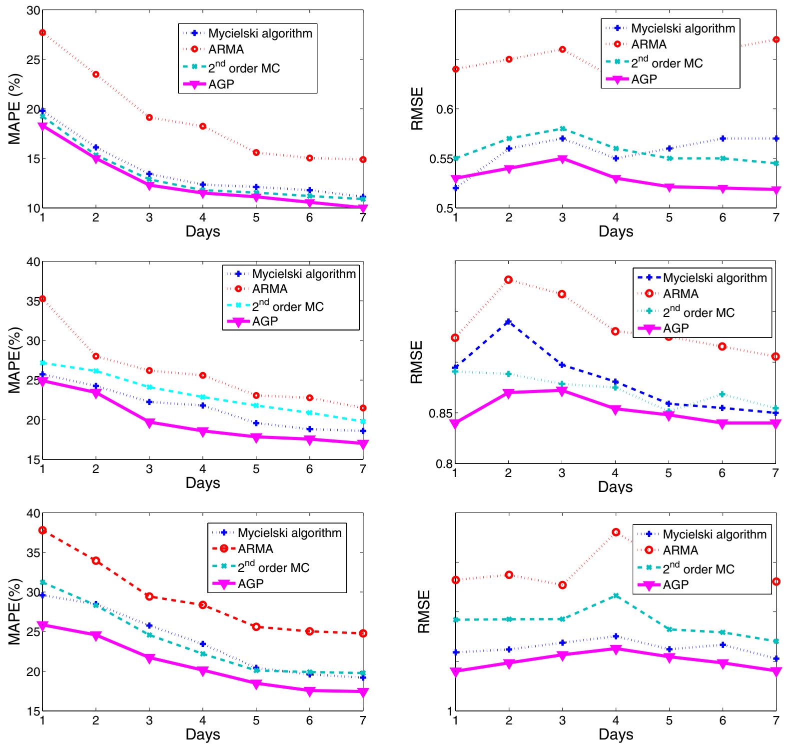
Figure 5a and 5b illustrate forecast results of different models for a 7-day period from 2/1/2007 to 2/7/2007 of the NY-ISO data and a 7-day period from 2/1/2010 to 2/7/2010 of the Calumet-ISO data, respectively. The results are depicted as markers on the graph at the end of each day, where aggregations of MAPE or RMSE are presented. AGP out-performances all the rest methods in both measurements. This shows adaptive Gaussian process has advantage in handling the high variability in wind speed time series. Note these results are data dependent and Figure 5 should not be read as MAPE or RMSE decreases over time.

Table 2: RMSE and MAPE evaluation of different methods. Note MAPE is a percentage of error while RMSE is a number indicating the model bias.

Method	**Range** (n-min ahead)	(NY-ISO)		(Calumet-ISO)	
		RMSE	**MAPE**(%)	**RMSE**	**MAPE**(%)
2^{nd} order MC	10	1.08	9.4	0.57	12.10
	20	1.50	14.78	0.98	19.05
	30	2.10	18.77	1.87	22.59
ARMA	10	1.32	12.16	0.69	14.82
	20	2.05	18.48	1.04	23.54
	30	2.53	22.77	2.61	27.84
Mycielski	10	0.96	9.79	0.58	12.37
	20	1.74	13.41	0.89	18.38
	30	1.97	17.19	1.63	22.07
AGP	10	0.65	9.02	0.54	10.2
	20	1.29	12.59	0.85	15.80
	30	1.60	15.43	1.41	17.36



(a) Results evaluation on NY-ISO dataset. The left column figure shows AGP outperforms our base lines in terms of MAPE and the right column illustrates our proposed method leads the RMSE performance. Note the X-axis indicates the total days included for comparison. Row 1,2 and 3 correspond to 10-minute, 20-minute and 30-minute ahead prediction, respectively.



(b) Results evaluation on Calumet-ISO dataset. The left column figure shows AGP outperforms our base lines in terms of MAPE and the right column illustrates our proposed method leads the RMSE performance. Note the X-axis indicates the total days included for comparison. Row 1,2 and 3 correspond to 6-minute, 12-minute and 18-minute ahead prediction, respectively.

Figure 5: Model Comparison, AGP, Mycielski algorithm, 2^{nd} order Markov Chain and ARMA are compared in terms of MAPE and RMSE.

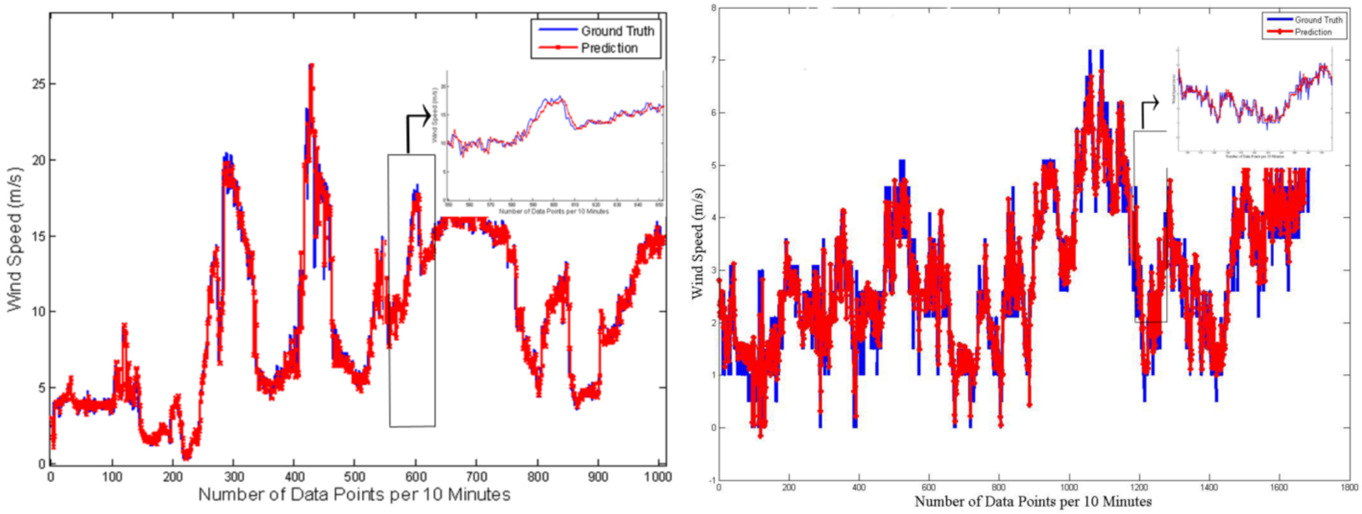


Figure 6: The graph shows AGP fitting against ground truth in NY-ISO (left column) and Calumet-ISO data (right column). Note the X-axis indicates the day of comparison. Y-axis represents the predicted value (red) and ground-truth value (blue).

Figure 6 shows the results of measured and predicted wind speed by AGP. We only plot points of 7 days for both datasets in order to make the figures uncluttered. Prediction results and ground-truth are shown in red and blue curves, respectively. Overall, the predicted wind speed based on AGP model closely follows the ground truth patterns and captures most of the local volatility. Please note that there some drastic changes in wind speed from 26m/s to 13m/s and predicted wind speed can track down to 15m/s. From the zoomed graph on the upper right corner, predicted wind speeds tracks well with measured on a daily base. From these consecutive forecasts illustrated, we observe the AGP has a strong descriptive power in handling different dynamic patterns in the wind speed.

Table 1 shows the detailed results of different methods applied on various forecasting time scale. From the results, we observe adaptive Gaussian process consistently outperforms other methods. For instance, in the 10-minutes ahead prediction of NY-ISO dataset, AGP method demonstrates significant improvments in RMSE and MAPE. For instace, RMSE performance has imporved103% against ARMA and 33% against Mycielski algorithm, meanwhile, MAPE has improved 8% and 4% against ARMA and Mycielski algorithm. For different time scale, the increase rates of RMSE and MAPE for different models are very similar from 43% (20-minute) to 120% (30-minute).

6 Conclusion

Renewable energy, especially wind energy, is recently very active topic. While traditional methods may not meet the specific requirements, this paper presents a novel model for short-term wind speed prediction for the purpose of better integration with electric energy system. We introduce an adaptive framework to handle the high variability in typical wind speed time series data. Three models have been compared with adaptive Gaussian process at different time scales. The results show that AGP outperforms other models in terms of RMSE and MAPE. Future research will focus on the real-time power grid controller design and implementation based on the predicted wind speed.

References

[1] *National Data Buoy Center, http://www.ndbc.noaa.gov/*, (2007).

[2] A.N. Celik, 'A statistical analysis of wind power density based on the weibull and rayleigh models at the southern region of turkey', *Renew Energy*, **29**, 593–604, (2004).

[3] Paul W. Goldberg, Christopher K. I. Williams, and Christopher M. Bishop, 'Regression with input-dependent noise: a gaussian process treatment', in *NIPS '97: Proceedings of the 1997 conference on Advances in neural information processing systems 10*, pp. 493–499, Cambridge, MA, USA, (1998). MIT Press.

[4] F. Hocaoglu, M.Fidan, and O. Gerek, 'Mycielski approach for wind speed prediction', *Energy Conversion and Management*, **50**, 1436–1443, (2009).

[5] Kristian Kersting, Christian Plagemann, Patrick Pfaff, and Wolfram Burgard, 'Most likely heteroscedastic gaussian process regression', in *ICML '07: Proceedings of the 24th international conference on Machine learning*, pp. 393–400, New York, NY, USA, (2007). ACM.

[6] K.Hunt and G.P. Nason, 'Wind speed modelling and short-term prediction using wavelets', *Wind Engineering*, **25**, 55–61, (2001).

[7] Q. V. Le, A. J. Smola, and S. Canu, 'Heteroscedastic gaussian process regression', in *ICML '05: Proceedings of the 22nd international conference on Machine learning*, pp. 489–496, New York, NY, USA, (2005). ACM.

[8] Carl Edward Rasmussen and Christopher K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*, The MIT Press, 2005.

[9] J.C. Palomares Salas, J.J. G. Rosa, J.G. Ramiro, J.Melgar, A. Aguera, and A. Moreno, 'Comparison of models for wind speed forecasting', Proc. of International Conference on Computational Science, (2009).

[10] Matthias Seeger, 'Gaussian processes for machine learning', *International Journal of Neural Systems*, **14**, 2004, (2004).

[11] A. Shamshad, M.A.Bawadi, WMA Wan Hussin, T.A.Majid, and SAM. Sanusi, 'First and second order markov chain models for synthetic generation of wind speed time m series', *Energy*, **30**, 693–708, (2005).

[12] S. Steftsos, 'A comparison of various forecasting techniques applied to mean hourly wind speed time series', *Renew Energy*, **21**, 23–35, (2000).

[13] M.C. Tore, P. Poggi, and A. Louche, 'Markovian model for studying wind speed time series in corsica', *Renew Energy*, **3**, 311–319, (2001).

[14] L. Xie and M. D. Ili, 'Model predictive economic/environmental dispatch of power systems with intermittent resources', Proceedings of IEEE Power and Energy Society General Meeting, (2009).

[15] F. Youcef, H. Sauvageot, and AHE Adane, 'Statistical bivariate modeling of wind using first order markov chain and weibull distribution', *Renew Energy*, **28**, 87–80, (2003).