Jurnal Teknik Industri, Vol. 22, No. 1, June 2020
ISSN 1411-2485 print / ISSN 2087-7439 online
DOI: 10.9744/jti.22.1.11-16

# Application of Ensemble Empirical Mode Decomposition based Support Vector Regression Model for Wind Power Prediction

**Irene Karijadi[1*], Ig. Jaka Mulyana[1]**

**Abstract**: Improving the accuracy of wind power prediction is important to maintain power system stability. However, wind power prediction is a difficult task due to non-stationary and high volatility characteristics. This study applies a hybrid algorithm that combines ensemble empirical mode decomposition (EEMD) and support vector regression (SVR) to develop a prediction model for wind power prediction. Ensemble empirical mode decomposition (EEMD) is employed to decompose original data into several Intrinsic Mode Functions (IMFs). Finally, a prediction model using support vector regression is built for each IMF individually, and the prediction result of all IMFs is combined to obtain an aggregated output of wind power. Numerical testing demonstrated that the proposed method could accurately predict the wind power in Belgium.

**Keywords**: Data mining, time series, prediction, renewable, energy.

## Introduction

Nowadays, electricity demand is increasing rapidly, and it cannot be supplied only from conventional energy sources such as fossil fuels since it has limited capacity. Therefore, there is a major shift in electricity generation where people started to utilize renewable energy technology, such as solar and wind energies. Compared with fossil fuels, renewable energy can reduce carbon emission and minimize the risk of electricity shortage. Wind energy is a part of renewable energy technology, and as reported by [1] the global cumulative installed wind capacity reached nearly 591 GW at the end of 2018.

Wind power prediction is essential to maintain power system stability. By having an accurate wind power prediction, utilities can adjust the power dispatching timely to ensure the stable operation of the power grid. However, the non-stationary pattern and strong volatility of wind speed characteristic make predicting wind power challenge.

Various methods have been developed to predict wind power, including statistical models and machine learning models. In [2] ARMA Statistical model is used to predict the tuple of wind speed and direction. Artificial Neural Network (ANN) as one of the machine learning method is utilized in [3] to predict wind dataset. Zhou *et al.* [4] utilized another popular machine learning method, namely Support Vector Regression (SVR) to predict one-step ahead of wind speed prediction and achieved reasonable accuracy level with fine-tuning SVR parameters.

[1]Faculty of Industrial Technology, Industrial Engineering Department, Universitas Katolik Widya Mandala, Surabaya, Jl. Kalijudan 37, Surabaya, 60236, Indonesia
Email: irenekarijadi92@gmail.com
* Corresponding author

Wind power data often shows a non-stationary pattern and high volatility; thus, it is difficult for a single method to predict the time series accurately. As a result, many researchers have developed a hybrid model to enhance the prediction accuracy of wind power prediction. Wavelet-based hybrid models have been proposed in [5], and wavelet could improve the accuracy of wind power prediction. However, the wavelet method needs a prior selection of basic wavelet and decomposition level. Empirical Mode Decomposition (EMD) is another decomposition approach introduced by Huang *et al.* [6]. One of the advantages of EMD over wavelet is its adaptiveness where EMD does not need a prior selection of decomposition level. The hybrid model incorporated EMD-SVR is introduced in [7]. EMD has a limitation of a mode mixing problem. Therefore Ensemble Empirical Mode Decomposition (EEMD) which is an improved version of EMD has been developed in [8] to tackle mode mixing problem.

EEMD decomposition method solves the mode mixing problem, one of the major drawbacks of the original EMD. Thus, in this paper, we applied EEMD-SVR to predict 15-minute ahead of wind power. Instead of EMD and wavelet decomposition, EEMD is employed in this study to decompose the original wind series data into a series of Intrinsic Mode Functions (IMFs) and one residue, thereby reducing the complexity of original wind series into relatively stationer subseries. SVR will be used to predict each Intrinsic Mode Functions (IMFs) and the residue. Predictions of each IMF and the residue component are aggregated by summation to obtain the final prediction.

**Table 1** EMD Algorithm

| **Algorithm 1** Empirical Mode Decomposition (EMD) Algorithm |
|---|
| **Step 1.** Given a time series S(t), identify local maxima and minima. |
| **Step 2.** Calculate upper $S_u[t]$ and lower $S_l[t]$ envelope by interpolation of local maxima and minima. |
| **Step 3.** Compute the mean ($m_t$)of upper and lower envelopes. $m_t = \frac{S_u[t]+S_l[t]}{2}$ |
| **Step 4.** Subtract mean from time series signal to obtain the first component h(t). $h(t) = S(t) - m_t$ |
| **Step 5.** Repeat the sifting process which consist of step 1 to step 4 by considering h(t) as new S(t) until one of the stopping criteria is reached:<br>(i) ($m_t$) approaches zero<br>(ii) the numbers of zero-crossings and extrema of h(t) differs at most by one, or<br>(iii) maximum number of iterations is reached.<br>The stopping criteria determines whether sifting process should stop to produce number of IMF |
| **Step 6.** Treat h(t) as new IMF and calculate the residual signal r(t) as $r(t) = S(t) - h(t)$ |
| **Step 7.** Regard the r(t) as the new time series S(t) to find next IMF. Repeat steps 1 to 6 until all IMFs are obtained |

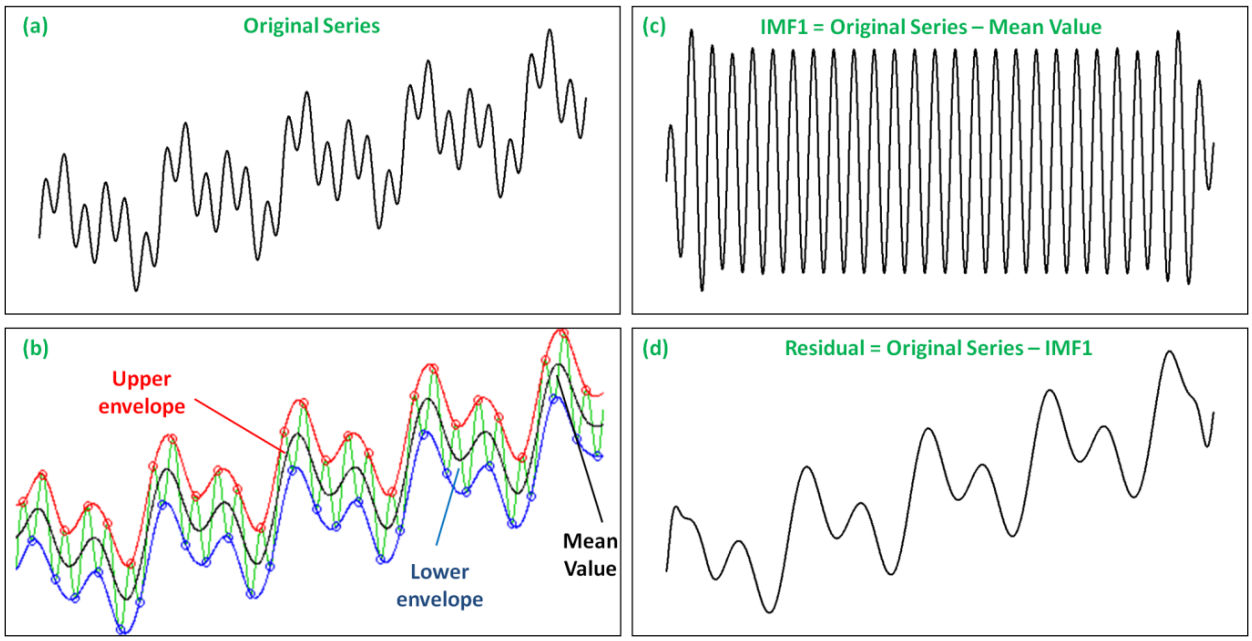


**Figure 1**. EMD decomposition illustration

## Methods

### Support Vector Regression

Support Vector Regression is a part of the data mining techniques. SVR aims to find a function that deviates from the actual observation obtained target by a value less than $\varepsilon$ for each training point [9]. A non-linear mapping function is used in SVR to project the training data set into a high dimensional feature space. The SVR attempts to minimize a minimization problem, as shown in the equation below:

$$\text{Min } \frac{1}{2}\|w^2\| \quad (1)$$

$$s.t \begin{cases} y_i - \langle w, x_i \rangle - b \le \varepsilon, \\ \langle w, x_i \rangle + b - y_i \le \varepsilon, \end{cases} \quad (2)$$

where $w$ is the vector of coefficients, $b$ is an intercept or bias term, and $x_i$ is a training sample with target value $y_i$. The inner product plus intercept $\langle w, x_i \rangle + b$ is the prediction for that sample and $\varepsilon$ is a free parameter that represents as a threshold.

### Ensemble Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) is capable of handling non-stationary and non-linear data, and it has been used in the past few years in the signal processing field and to improve the prediction accuracy. EMD reconstructs a time-series signal into a set of Intrinsic Mode Functions (IMFs) along with a residual trend. The procedure of the EMD algorithm [10] is described in Table 1.

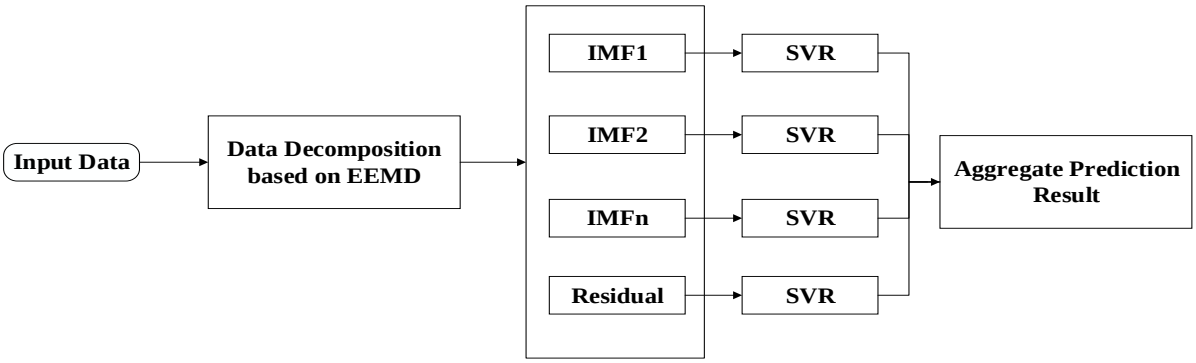


**Figure 2**. EEMD-SVR Flowcharts

Figure 1 illustrates the idea of EMD decomposition from step 1 to step 6, where Fig.1.a shows the original series and Fig.1.b shows its upper (red lines) and lower envelopes (blue lines), as well as the mean of the envelopes (black lines). The first IMF obtained from subtracting the mean value from the original time series (Fig.1.c), and Residue obtained from subtracting the first IMF from the original time series (Fig.1.d).

Ensemble Empirical Mode Decomposition (EEMD) is a recent improvement from the original author of EMD in which additional noise is used to better separate different frequency scales into different IMF. EEMD procedure is as follows [8]: (1) Add Gaussian white noise to input data. (2) Decompose the data with added white noise into IMFs. (3) Repeat step 1 and 2 by $N$ times. $N$ is the number of ensembles. Gaussian white noise with difference mean and variance is added for each time. (4) Calculate the means of corresponding IMFs as the final decomposition result.

## Methodology

The proposed methodology, which comprises of support vector regression (SVR) and ensemble empirical mode decomposition (EEMD) for wind power prediction, is illustrated in Figure 2. The proposed methodology consists of two stages, which are data decomposition and data prediction.

### *Data Decomposition*

The wind power data often shows a non-stationary pattern and high volatility. To reduce the effect of volatility and non-stationary pattern, the original time series data will be decomposed using Ensemble Empirical Mode Decomposition (EEMD) into several numbers of subseries (n). The number of subseries is determined from the sifting process, and this subseries can be respectively named as IMF1, .,.., IMFn.

**Figure 3.** Wind turbine in Belgium [15]

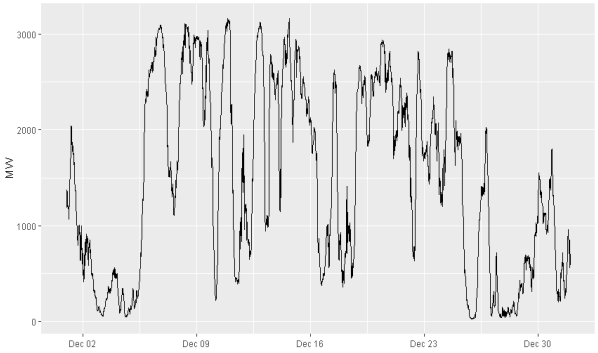


**Figure 4.** Wind power generation over one-month period

### *Data Prediction*

After that, the SVR prediction model is built for each IMF and residual. In this study, the input for the prediction model includes the previous half-hour of wind power (2 as the time-lagged). After building SVR for each component, the prediction output of every IMF and residual are combined by summation as depicted in Figure 4.

## Results and Discussions

### Experimental Result

We compared the hybrid SVR performances for each building with traditional SVR model, linear regression, random forest model and EMD-SVR. We applied

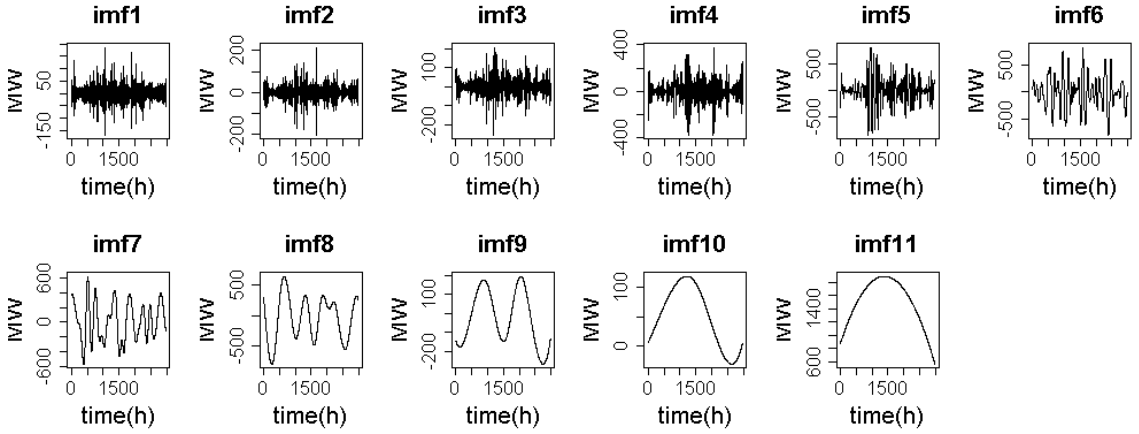


**Figure 5.** Decomposition result

EEMD using the "Rlibeemd" package in [11], and for SVR prediction model, we used the "e1071" package in R [12]. All the experimental results were conducted using R programming on a standard PC.

## Dataset

The proposed method was tested on a publicly available dataset for wind power generation in a Belgian wind farm [13]. Figure 3 shows one of the wind turbines in Belgium. Wind turbines are utilized to capture kinetic energy from the wind and generate electricity. When the wind blows past a wind turbine, its blades capture the kinetic energy from the wind and rotate it, turning it into mechanical energy. The rotation itself turns an internal shaft connected to a gearbox, which increases the speed of rotation and that spins a generator that produces electricity [14]. The period of data used in this study is from 1 December 2019 to 31 December 2019. Figure 4 shows the original time series data of wind power generation for one-month period [15]. Data were collected with a time interval of 15 min.

## Prediction Model Development

### *Input*

We built the SVR prediction model for each IMF and residual. For 15 minutes ahead of wind power prediction ($X_t$), the input feature includes the previous 30 min wind power data ($X_{t-1}$ , $X_{t-2}$)

### *Parameter Setting*

SVR has several parameters that are required to be determined in advance. In this study, we used a grid search for hyper-parameter tuning. We used RBF function as our kernel function. The range of cost is [2-4,24], and the range of gamma is [10-3,10-1].

### *Training and Testing*

We divided the data into 70% of data as training data set and 30% of data as test data sets. The training data set is used to develop the models while the test data set is used to evaluate the prediction performance of the models.

## Evaluation Metrics

We use Mean absolute percentage error (MAPE), root means square error (RMSE), and mean absolute error (MAE) as evaluation metrics to verify the performance of the proposed method. The formula for MAPE, RMSE and MAE are given in equations (4–6) below:

$$MAPE\ \% = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t' - y_t}{y_t}\right| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t' - y_t)^2} \quad (5)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t' - y_t| \quad (6)$$

where $y_t'$is the predicted value, $y_t$is the actual value, and $n$ is the number of data points in the time series. Since MAPE, RMSE and MAE is a measure of error, high numbers are bad and low numbers are good.

## The Results

In this study, the EEMD method is employed with 250 as the ensemble number. EEMD reconstructs the original time series into several IMFs and one residue. Figure 5 displays the decomposition results of original time series data. EEMD decomposed into 11 subseries which this 11 subseries can be respectively named as IMF1, IMF2,.., IMF11. The highest frequency series is IMF1, and IMF11 is the lowest frequency among these series which also reflects the trend of the original series. IMF11 is also named as residue. As can be seen from Figure 5, the extracted component obtained from EEMD is more stable than the original data, and the extracted components are easier to be modelled.

**Table 2**. Sample entropy value for each IMF

| IMF 1 | IMF 2 | IMF 3 | IMF 4 | IMF 5 | IMF 6 |
|---|---|---|---|---|---|
| 0.946 | 0.835 | 0.761 | 0.683 | 0.558 | 0.489 |
| IMF 7 | IMF 8 | IMF 9 | IMF 10 | IMF 11 | |
| 0.437 | 0.354 | 0.313 | 0.284 | 0.286 | |

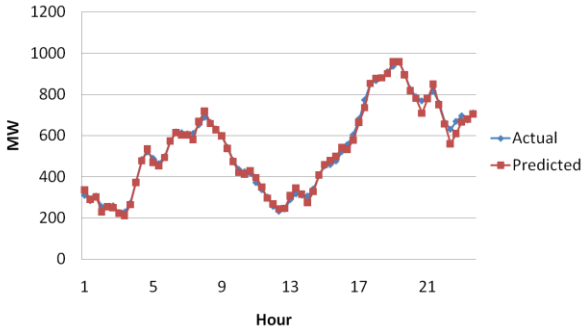


**Figure 6.** Prediction result on test dataset (1 day observation)

Shannon entropy (SE) [16] is used to measure the volatility of the series. High values of this statistic indicate volatility and unpredictability. Table 2 shows the SE of the sub-series, and the first IMF has the biggest SE, and as the number of IMF increases, the SE of each IMF is greatly reduced.

After being decomposed by EEMD, we build the SVR model on each IMF and residue. After building SVR for each component, the prediction output of every IMF and residual are combined by summation.

The predicted results of EEMD-SVR model (red lines) fit and are close to the actual data (blue lines), which demonstrates that the EEMD-SVR can produce good prediction result to predict wind power (see Figure 6).

We benchmarked the proposed method with other machine learning methods such as linear regression, random forest (RF), support vector regression (SVR) and Empirical mode decomposition combined with support vector regression (EMD-SVR) [7]. For all the machine learning methods (Linear Regression, Random Forest, SVR, EMD SVR, and EEMD SVR), the output is the 15 minutes ahead wind power prediction, and the input feature includes the previous 30 minutes of wind power data. For Random Forest (RF), we used the "e1071" package in R [12], and for Linear Regression, we used the "stats" package in R [17]

The results of the proposed method are summarized in Table 3. As presented in Table 3, we can see that the proposed method has outperformed the benchmark method. Besides that, compared with SVR, the prediction accuracy after the application of EEMD prior to SVR is greatly enhanced. Thus, we can conclude that EEMD is an effective preprocessor to improve prediction accuracy and a proper technique to handle data with non-stationary and volatility.

## Conclusion

An accurate and reliable wind power prediction is crucial to reduce the operating cost of wind power and to maintain grid stability. However, it is often difficult to predict wind power accurately because wind power has the characteristic of high volatility and non-stationary. In this paper, a hybrid EEMD based Support Vector Regression is applied to predict the wind power in one of the wind farms in Belgium. The experimental results indicate that the proposed method produces better results compared to the traditional SVR method, linear regression, random forest, and EMD SVR method. The EEMD is employed as data pre-processing to transform the original data into more stable subseries, and from the experimental results, we can see that EEMD is a good decomposition strategy to enhance the prediction accuracy. For future work, some external factors, such as weather data will be incorporated into the model.

## References

1. Council, G.W., Global Wind Energy Report: Annual Market Update 2017, *Global Wind Energy Council: Bruxelles,* Belgium, 2018.
2. Erdem, E., and Shi J., ARMA based Approaches for Forecasting the Tuple of Wind Speed and Direction, *Applied Energy,* 88(4), 2011, pp.1405-1414.
3. Li, G., and Shi J., On Comparing Three Artificial Neural Networks for Wind Speed Forecasting, *Applied Energy,* 87(7), 2010, pp.2313-2320.
4. Zhou, J., Shi, J., and Li, G., Fine Tuning Support Vector Machines for Short-term Wind Speed

**Table 3.** Summary of prediction results

| Method | Training | | | Testing | | |
|---|---|---|---|---|---|---|
| | MAPE (%) | RMSE | MAE | MAPE (%) | RMSE | MAE |
| Linear Regression | 4.103544 | 67.72832 | 44.57145 | 6.957862 | 49.59483 | 35.46429 |
| Random Forest | 2.225683 | 35.8316 | 24.01349 | 10.11488 | 58.53541 | 42.86142 |
| SVR | 3.98134 | 67.78834 | 44.44633 | 6.342731 | 49.43159 | 35.13028 |
| EMD SVR | 2. 152155 | 37.27049 | 24.33087 | 5.996273 | 34.28776 | 21.80686 |
| EEMD SVR | 1.693411 | 20.79005 | 14.84377 | 3.872025 | 16.83474 | 12.85613 |

Forecasting, *Energy Conversion and Management*, 52(4), 2011, pp. 1990-1998.
5. Liu, H., Tian, H.Q., Pan, D.F., and Li, Y.F., Forecasting Models for Wind Speed using Wavelet, Wavelet Packet, Time Series and Artificial Neural Networks, *Applied Energy*, 107, 2013, pp. 191-208.
6. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., and Liu, H.H., The Empirical Mode Decomposition and the Hilbert Spectrum forNonlinear and Non-stationary Time Series Analysis, *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971), 1998, pp. 903-995.
7. Zhang, W., Liu, F., Zheng, X., and Li, Y., A Hybrid EMD-SVM based Short-term Wind Power Forecasting Model. *In 2015 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC),* 2015, pp. 1-5.
8. Wu, Z., and Huang, N.E., Ensemble Empirical Mode Decomposition: A Noise-assisted Data Analysis Method, *Advances in Adaptive Data Analysis*, 1(01), 2009, pp. 1-41.
9. Smola, A. J., and Schölkopf, B., A Tutorial on Support Vector Regression, *Statistics and Computing*, 1;14(3), 2004, pp. 199-222.
10. Maheshwari, S., and Kumar, A., Empirical Mode Decomposition: Theory and Applications, *Journals of Electron and Electrical Engineering*, 7, 2014, pp. 873-8.
11. Helske, J., Luukko, P., and Rlibeemd: Ensemble Empirical Mode Decomposition (EEMD) and Its Complete Variant (CEEMDAN), *R package version*,1(0), 2017.
12. Meyer, D., Dimitriadou, E., and Hornik, K., Weingessel A, Leisch F., e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, 2017, *R package version,* 1(8), 2018.
13. Elia NV., *Wind-Power Generation Data*, retrieved from https://www.elia.be/en/grid-data/power-generation/wind-power-generation
14. *Basics of Wind Energy*. retrieved from https://www.awea.org/wind-101/basics-of-wind-energy
15. *Wind Energy in Belgium*, retrieved from https://community.ieawind.org/about/member-activities/belgium
16. Sheraz, M., Dedu, S., and Preda, V. Entropy Measures for Assessing Volatile Markets. *Procedia Economics and Finance, 22,* 2015, pp. 655-662.
17. Team RC., *R: A Language and Environment for Statistical Computing*, 2013