PROCEEDINGS OF SCIENCE

Short-term Wind Speed Forecasting by Using Model Structure Selection and Manifold Algorithm

Haijian Shao[1]
School of Automation, Southeast University, Jiangsu, 210096, Nanjing, China
Key Laboratory of Measurement and Control for Complex System of Ministry of Education, Southeast University, Jiangsu, 210096, Nanjing, China
E-mail: haijianshao@seu.edu.cn

Xing Deng
School of Automation, Southeast University, Jiangsu, 210096, Nanjing, China
Key Laboratory of Measurement and Control for Complex System of Ministry of Education, Southeast University, Jiangsu, 210096, Nanjing, China
E-mail: xingdeng@seu.edu.cn

Nonlinear model structure selection (MSS) is an important step in the modeling theory of the nonlinear system. The proper model structure and model parameters estimation can be used to reduce unnecessary computations and improve the model forecast accuracy. This paper mainly investigates the short-term wind speed forecasting (STWSF) by utilizing the comprehensive MSS technique based on real data of the wind speed plant in East China. The MSS and manifold algorithm are respectively used to design proper model structure and reduce the computational complexity to improve the forecasting accuracy and promote the computational efficiency. The experimental evaluation by using support vector regression based on the real data is given to demonstrate the performance of the proposed method.

ISCC 2015
18-19, December, 2015
Guangzhou, China

[1]Speaker

This project is supported by the National Natural Science Foundation of China (No. 61374006).

PoS(ISCC2015)027

1. Introduction

The distribution of wind speed is volatile, intermittent and random so that its model structure selection (MSS) is relatively difficult. The wind speed forecasting has significant impact for security and stability of the large-capacity wind speed connected to the grid. Most forecasting methods such as time series[1], Kalman filtering[2], Artificial neural networks (ANNs)[3], numerical weather prediction[4], fuzzy logic method[5] and support vector machine[6], don't take into account the distribution characteristics of meteorological data. The main disadvantages of the outlined literature include that: they do not provide a complex analysis for the model structure design and selection, such as the model variables selection, model variable order estimation as well as the performance measurement criteria, typically, there are plenty of available variables which can be used for modeling; however, the proper variables with approximately model order are not provided, which results in lacking the ability of generalization and practicability.

Manifold algorithm (MA) is a mature technology, which can be applied for the input variable dimensionality reduction, automated target recognition[7] and machine intelligence [8] etc. Bloch used MA to solve the conflictive situations throughout the information combination based on the classification behavior [9]. Chou et al. [10] introduced the multi-scale recursive estimation and MA for a traditional dynamical model with dyadic trees. Banerjee et al. [11] used the sparse representation and MA for the MSS based on the maximum likelihood estimation and implemented the memory savings methods even if the utilized the method contains more than tens of nodes. Dietterich [12] investigated five approximate statistical tests for the learning algorithm according to different types of errors, and then made the proper decision for the MSS based on the cross-validation and MA. Although the outlined methods have proposed many useful strategies for the practical issue, the utilized model is usually not a robust and reasonable method because the model structure design basically depends on the subjective experience, for instance, the dynamic characteristics of the system is not considered. In addition, the traditional forecasting method of the numerical weather prediction is instabile because the climate of different wind speed plants is different. In addition, the selected model is not very used for the variables with large dimension due to high computational complexity.

Based on the above discussion, the accurate and appropriate models can be used to reflect the characteristics of the utilized variable, and the sample feature extraction benefit the improvement of the computational efficiency. How to select proper model structure, reduce the computational complexity and improve the forecast accuracy based on the characteristics of samples is what we want to address in this paper.

2. Data Description

The power generation equipment used weak wind turbine type wind turbines--FD77 of Dongfang Steam Turbine Co., Ltd., the diameter of wind wheel is 77m, and wheel height has 3 different heights, respectively, 57.65m, 66.25m and 81.25m. In this Section, the utilized data is from 2011/06/01 to 2012/06/01 and the sampling frequency is 5 minutes. Based on different heights 57.65m, 66.25m and 81.25m of the wheel height. 18 input variables are divided into four groups based on different heights (10m, 50m and 70m):

Group 1 (G1). 10m: No1, AWS; No2, AWD; No3, SSD; No4, RTWS; No5, RTWD;

Group 2 (G2). 50m: No6, AWS; No7, AWD; No8, SSD; No9, RTWS; No10, RTWD;

Group 3 (G3). 70m: No11, AWS; No12, AWD; No13, SSD; No14, RTWS; No15, RTWD;

Group 4 (G4). No16, 10m temperature; No17, 10m humidity; No18, Pressure.

where AWS: average wind speed; AWD: average wind direction; SSD: sample standard deviation; RTWS: real-time wind speed; RTWD: real-time wind direction.

3. Model Structure Selection

3.1 Model Variable Selection

Model Variable Selection (MVS) is an essential step for the modeling with plurality of available variables. The proper variables applied as inputs should be highly penetrated information system with appropriate number, independent with each other and used for the output representation. Taking the neural network as an example, MIV can be used to measure the weight matrix changes of the neural network to determine the impact selection index related to the output and input neurons. The independent variable feature of the relevant sample by increasing and decreasing 10% (empirical values proposed by the original author) respectively to formulate the new training sample set ISI1 and ISI2 which are defined by

$$\{x_{i,j}(t)\}_{i,j} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} \pm r \cdot x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} \pm r \cdot x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{k-1,1} & x_{k-1,2} & \cdots & x_{k-1,j} \pm r \cdot x_{k-1,j} & \cdots & x_{k-1,n} \\ x_{k,1} & x_{k,2} & \cdots & x_{k,j} \pm r \cdot x_{k,j} & \cdots & x_{k,n} \end{bmatrix} \tag{3.1}$$

$i = 1,...,k, j = 1,...,n$. IS1 and IS2 are respectively used as the new input variables for the neural network testing, then the impact value about the input variable to the output variable is derived by the difference between S1 and IS1, S2 and IS2, respectively.

3.2 Model Order Estimation

Usually, the model variable order has the essential connections associated with the current output variables, which reflects the dynamical system's maximum time-delay. Typically, the larger the system's dynamics persistence is, and then the higher the model variable order will be.

4. Data Dimensionality Reduction

4.1 Manifold Algorithm

Many methods had been developed for that issue in recent years although these methods came from many different disciplines [13]. Many well-known methods, such as the principal components analysis (PCA) [14], the multidimensional scaling (MDS) [15], the independent component analysis (ICA) [16] and the factor analysis [17], can model the linear subspace (manifolds) for a given high dimensional observation; however, these methods have one common drawback: they only focus on the data characteristics in linear subspace; therefore, many excellent methods, such as kernel PCA [18,19], locally linear embedding (LLE) [20,21], Laplacian eigenmaps (LEM) [22], Hessian eigenmaps (Hessian LLE) [23], locality preserving projections (LPP) [24] and Landmark isomap [25] can be used to solve the dimensionality reduction problem in the nonlinear space.

4.2 Intrinsic Dimensionality Estimation

Christopher [26] figured out that the intrinsic dimension can be defined via a scale-dependent quantity method. For instance, the Nystrom method is a typically technique used for the numerical approximation in the manifold algorithm. In this section, the correlation dimension is used for the intrinsic dimension estimation, which is similar to the fractal dimensions used in fractal geometry. If the finite set $\zeta_n = \{x_1, x_2, ..., x_n\}$ in metric space,

$$C_n(r) = \alpha \sum_{i=1,\dots,n} \sum_{j=i+1,\dots,n} I_r, \quad I_r = \left\{ \|x_i - x_j\|^2 < r \right\}, \tag{4.1}$$

where $\alpha = 2 \times (n(n-1))^{-1}, I_r = \left\{ \|x_i - x_j\|^2 < r \right\}$, is the corresponding index set. The integral is given by $C(r) = \lim_{r \ 0} C_n(r)$ for a countable subset $S = \{x_1, x_2, ...\}$ X . If $C(r)$ can be derived, then the correlation dimension of ζ_n is defined by $CD_{corr} = \lim_{r \ 0} \frac{\log C(r)}{\log r}$. Assume the distribution of the data at the high-dimensional manifold is uniform, the corresponding intrinsic correlation dimension is defined by

$$C\hat{D}_{corr} = \frac{\log C(r_2) - \log C(r_1)}{\log r_2 - \log r_1} \tag{4.2}$$

Thus, the intrinsic dimension is estimated before the nonlinear mapping is established.

5. Support Vector Regression and Performance Criteria

5.1 Support vector regression

The mainly task of SVM for regression is to give the nonlinear function

$$y = f(x)\big|_{R^N \to R} = \sum_{i=1}^{N} \omega_i \varphi_i(x) + b \tag{5.1}$$

So that the variable x could be mapped into a higher dimensional feature space related to the samples set $S = \{(x_i, y_i)\}_{i=1}^{N}$, where $\{\omega_i\}_{i=1}^{N}$ and b are the coefficients to be calculated by (5.1), $\{\varphi_i(x)\}_{i=1}^{N}$ is the feature set. The nonlinear regression is mapped into a linear regression via a lower dimension input space to a higher dimensional space.

$$\min R(\omega) = \lambda \|\omega\|^2 + C \sum_{i=1}^{N} (\eta_i^* + \eta_i),$$
$$s.t. \quad \begin{cases} y - k \sum_{i=1}^{N} \omega_i \varphi_i(x) - b \le \varepsilon + \eta_i^*, \\ k \sum_{i=1}^{N} \omega_i \varphi_i(x) + b \le \varepsilon + \eta_i + y, \quad \eta_i^*, \eta_i \ge 0 \end{cases} \tag{5.2}$$

where "*s. t.*" denotes the "subject to", C is a penalty factor, λ is a regularization constant, η_i^*, η_i are relaxation factors.

5.2 Performance Criteria

The fitting performance indicators are measured by the following formulas,

$$RMSE = \frac{1}{\sqrt{n}} \times \sqrt{\sum_{t=1}^{n} \left(W_{tf} - W_{tr}\right)^2}, RMAE = \frac{1}{\sum_{t=1}^{n} \left|W_{tr}\right|} \times \sum_{t=1}^{n} \left|W_{tf} - W_{tr}\right| \tag{5.3}$$

where RMSE and RMAE are respectively the root mean square error and relative average absolute error associated to forecasting data W_{tf} and real data W_{tr}, n is the length of test sample.

6. Experimental Evaluation

In order to illustrate clearly the main proposed approach of this paper, the following flow chart is given to demonstrate each processing steps of this paper.

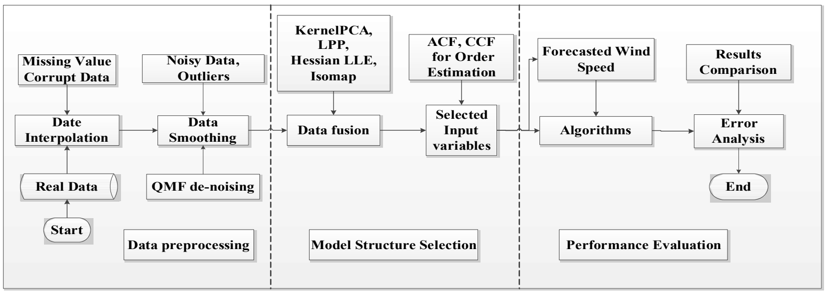

Figure 1: Main Processing Block.

6.1 Data Preprocessing

Daubechies wavelet, and the parameter related to the support and vanishing moments of the wavelets set 3. The parameter for low frequency cutoff for shrinkage is 5, and the noise estimation level is normalized to be 1. State trajectories of part of original variables and filtered variables after interpolation by QMF are shown in Fig. 2

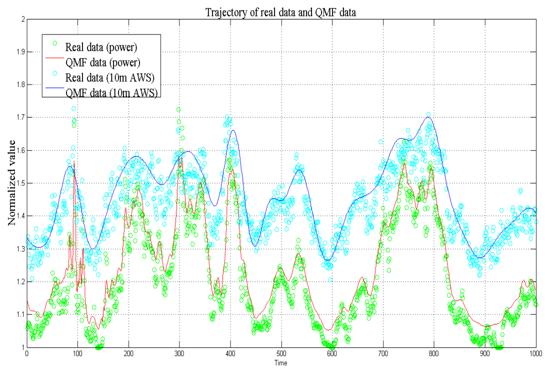

Figure 2: Trajectory of 10m AWD and Corresponding Power.

Trajectories of variables regard to 10m AWS and corresponding power are selected to be displayed in Fig. 2, respectively.

6.2 Model Variable Selection and Order Estimation

The main purpose of the MVS is to select the proper input variable for the modeling, so the MIV is calculated based on the real data from a wind farm of East China. 70m AVS is selected as the output variable because it is the nearest observation point in the wind tower and the influence from the ground friction is small. Our previous work [3] indicated that the correlation of the same variable with different heights in different groups was relative large, such as the correlation coefficient between the wind speed at different heights and the corresponding standard deviation was large while the correlation between the wind speed and temperature or humidity was weak. It is noted that the cosine and sine of wind direction should be selected as the input variable at the same time because the wind direction around zero angle is generally the same with each other. The obtained mean impact value about the Model Variable Selection (MVS_MIV) is listed as Table 1.

MIV	Six months	Nine months
x_1	0.0039	0.0050
$cos(x_2)$	-0.0048	-0.0065
$sin(x_2)$	-0.0050	-0.0059
x_3	0.0012	0.0015
x_4	0.0032	0.0045
$cos(x_5)$	-0.0050	-0.0061
$sin(x_5)$	-0.0048	-0.0061
x_6	**0.0074**	0.0060
$cos(x_7)$	-0.0050	-0.0060
$sin(x_7)$	-0.0048	-0.0062
x_8	7.6983e-04	0.0030
x_9	0.0068	0.0060
$cos(x_{10})$	-0.0050	-0.0060
$sin(x_{10})$	-0.0049	-0.0063
x_{11}	**0.0080**	0.0062
$cos(x_{12})$	-0.0052	-0.0062
$sin(x_{12})$	-0.0048	-0.0057
x_{13}	5.3146e-05	0.0033
x_{14}	**0.0077**	0.0062
$cos(x_{15})$	-0.0051	-0.0062
$sin(x_{15})$	-0.0049	-0.0060
x_{16}	-0.0031	-0.0054
x_{17}	-0.0029	-0.0057
x_{18}	-0.0044	-0.0056

Table 1: MVS_MIV Wind Farm of East China

The calculation results about the MVS_MIV are derived based on the RBF neural network, the convergence goal is 0.001 and the spread selection utilized the default value. The changes of the neural network weight matrix can be used to measure whether the input variable has potential influence to the outputs. The new training sample is obtained by increasing or decreasing the original input variable by 10% then used to calculate the impact value based on the difference between the original inputs and new ones. Sine and cosine of the wind direction is simultaneously for the input variable in order to reflect its characteristics fully. Based on Table 1, 50m AWS, 70m AWS and RTWS have relatively larger MVS_MIV than other available variables, the value of which remains more than 0.0074. This also indicates that the previous correlation analysis about the available input variables is correct. It is noting that the absolute value of the MVS_MIV (nor positive and negative original value) is the index to measure if the input variable has potential influence for the outputs.

6.3 Model Order Estimation

Essentially, the most influential variable for the power is wind speed, for instance, the absolute value of correlation between power and G2: No.6 (AWS), G3: No.11 (AWS) and No14 (RTWS) are more than 0.9; in this sense, in order to investigate the best input variable for the power forecasting, each wind will be forecasted and used to predict the power; moreover, different combinations of variables will be selected as the input variable then the best variables for the wind forecasting will be derived.

6.4 Intrinsic Dimension Estimation

The intrinsic dimension estimation is critical for the input variable dimensionality reduction because the improper dimension cannot reflect the variable information accuracy. The dimensionality estimation is calculated according to four different information criteria and listed in Table 2.

Data Length	CorrD	NearND	MLE	GMST
1000	1	1	1	2
6000	1	1	1	2
9000	1	1	1	2

Table 2: Dimensionality Estimation

where CorrD represents the computing method in the section of Intrinsic Dimensionality Estimation, NearND, MLE and GMST represents the nearest neighbor dimension, maximum likelihood estimator and Geodesic minimum spanning tree, receptively.

6.5 Model Variables Dimensionality Reduction

Without loss of the generality, the Kernel principal component analysis (KPCA), isomap, Locally Linear Embedding (LLE) and Local Preserving Projection (LPP) techniques are used to the corresponding data fusion of AWS and RTWS. Trajectories of 10AWS, 50AWS and 70AWS with corresponding data fusion are given in Fig. 3

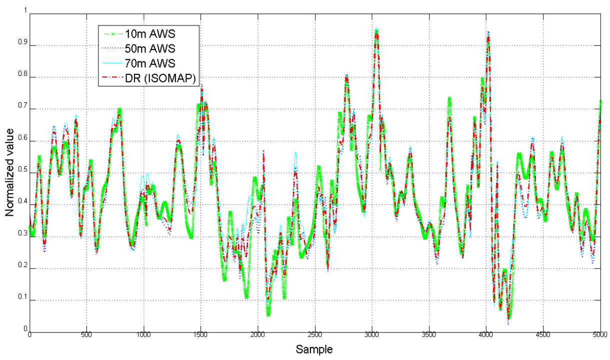

Figure 3: Dimensionality Reduction via Isomap for AWS.

As the number of nearest neighbors in a neighborhood graph is 3, the size of new variable based on the data reduction technique is one third compared to the original data. The elapsed times with respect to Isomap, KPCA, LLE and LPP are given in Table 3,

Item	Isomap	KPCA	LLE	LPP
AWS	8621.2727	965.5666	52.4120	45.0961
AWD	8580.9896	1504.7523	59.5469	48.0975
SSD	8795.1781	961.1593	56.9572	55.2840
RTWS	8850.2988	1022.3562	62.3404	55.3235
RTWD	8698.6513	965.3235	62.1589	50.9132

Table 3: Elapsed time in Seconds

6.6 Error Analysis and Evaluation

In this section, SVR technique is used to verify the performance of power forecasting. "Numerical meshgrid search" method is used to optimize the parameters of kernel function, which is to try every possible parameters pairs (c, g) values, where c and g are respectively penalty factor and kernel function parameters. 60% (54760/91226), 20% and 20% of each subset are respectively defined as the training, verification and testing sample about 12 steps

PoS(ISCC2015)027

ahead predict (i.e. 1h). STWSF using MSS in combination with four methods of DR are given in Table 4 and Fig.4.

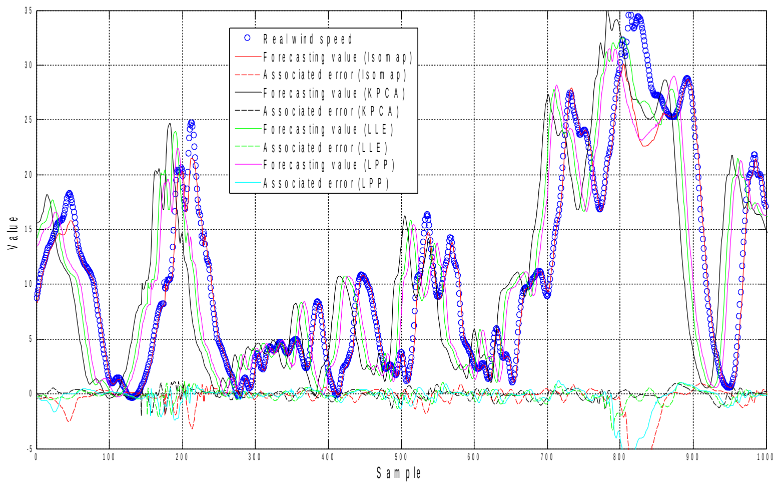

Figure 4: STWSF Using Four DR Methods.

Item	BC	BG	RMSE	RMAE	ET
TRA	0.8362	3.20	1.7303	0.2506	2848e+04
Isomap	0.3125	3.20	1.5469	0.2096	2330.32
KPCA	0.6250	3.15	0.6598	0.1750	2220.25
LLE	0.3165	3.20	0.7294	0.1779	2144.22
LPP	0.2265	3.80	1.7916	0.2162	2070.32

Table 4: Error Analysis and Parameters Estimation

BC: Best C; BG: Best G; TRA: traditional approach with no MSS; RMSE: root mean square error; RMAE: relative average absolute error; MSS-KPCA: kernel principal component analysis; MSS-LLE: Locally Linear Embedding; MSS-LPP: Local Preserving Projection; ET: Elapsed time for testing sample in seconds. 1000*9 data (about 3.5 days) for 24 steps-ahead prediction is shown in Fig. 6. As a result, we can draw a conclusion as follows:

(1) MSS has significant impact for the forecasting accuracy because it can not only select significant variables for STWSF but also reduce the dimensionality, the computational complexity while increase the prediction efficiency.

(2) The forecasting accuracy of KPCA is general better than other data fusion techniques and its forecasting accuracy is improved by about 11.74%, 1.54% and 14.43% respectively.

7. Conclusion

As the weather fluctuations as well as the seasonal changes can affect the power forecasting accuracy, accurate forecasting will improve the security and stability of power systems. MSS has significant reference value for the STWSF. Firstly, in order to guarantee the quality of the data for the further analysis, the missing value, noisy data and extreme are handled by the numerical interpolation, QMF and normalization, respectively; secondly, MSS which constitutes of model variable selection and order estimation is proposed in order to derive the model with proper structure; thirdly, the input variables dimensionality reduction is implemented via four different manifold techniques to obtain MSS with high computational efficiency; finally, the performance evaluation of MSS and error analysis is given to illustrate that the proposed method is a successful method for STWSF.

References

[1] Gang, Ding, Zhong Shi-Sheng, and Li Yang. *Time series prediction using wavelet process neural network.* CHINESE PHYS B 2008, 6(6), 1998-2003.

[2] Petroula Louka, Georges Galanis, Nils Siebert, Georges Kariniotakis, Petros Katsafados. *Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering.* J WIND ENG IND AEROD 2008, 96(12), 2348-2362.

[3] Wei, Haikun, Shao Haijian, and Deng Xing. *Using a Model Structure Selection Technique to Forecast Short-Term Wind Speed for a Wind Power Plant in North China*. J ENERG ENG 2015, 4015005.

[4] Hu Shu-Juan, Qiu Chun-Yu, Zhang Li-Yun, Huang Qi-Can, Yu Hai-Peng and Chou Ji-Fan. *An approach to estimating and extrapolating model error based on inverse problem methods: towards accurate numerical weather prediction.* CHINESE PHYS B 2014,08(8):669-677.

[5] Kukolj, Dragan, and Emil Levi. *Identification of complex systems based on neural and Takagi-Sugeno fuzzy model.* Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 2004, 34(1), 272-282.

[6] Shen Yang-Wu, Ke De-Ping, Sun Yuan-Zhang, Daniel Kirschen, Wang Yi-Shen and Hu Yuan-Chao. *A novel transient rotor current control scheme of a doubly-fed induction generator equipped with superconducting magnetic energy storage for voltage and frequency support.* CHINESE PHYS B 2015, 24, 70201.

[7] Amani Tahata, Jordi Martia, Ali Khwaldehb, and Kaher Tahatc. *Pattern recognition and data mining software based on artificial neural networks applied to proton transfer in aqueous environments.* CHINESE PHYS B 2014, 23, 46101.

[8] Abidi, Mongi A., and Rafael C. Gonzalez. *Data fusion in robotics and machine intelligence.* Academic Press Professional, Inc.: 1992;

[9] Bloch, Isabelle. *Information combination operators for data fusion: a comparative review with classification.* Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 1996, 26(1), 52-67.

[10] Chou, Kenneth C., Alan S. Willsky, and Albert Benveniste. *Multiscale recursive estimation, data fusion, and regularization.* Automatic Control, IEEE Transactions on 1994, 39(3), 464-478.

[11] Banerjee, Onureena, Laurent El Ghaoui, and Alexandre d'Aspremont. *Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data.* The Journal of Machine Learning Research 2008, 9, 485-516.

[12] Dietterich, Thomas G. Approximate statistical tests for comparing supervised classification learning algorithms. NEURAL COMPUT 1998, 10(7), 1895-1923.

[13] Ghodsi, Ali. *Dimensionality reduction a short tutorial*. Department of Statistics and Actuarial Science, Univ. of Waterloo, Ontario, Canada 2006.

[14] Jolliffe, Ian. *Principal component analysis*. Wiley Online Library: 2005;

[15] Cox, Trevor F., and Michael AA Cox. *Multidimensional scaling*. CRC Press: 2010;

[16] Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning.* Springer Series in Statistics New York: 2001;

[17] Frey, Brendan J. *Graphical models for machine learning and digital communication*. MIT press: 1998

[18] Mika, S., Schölkopf, B., Smola, A. J., Müller, K. R., Scholz, M., & Rätsch, G. *Kernel PCA and De-Noising in Feature Spaces*., NIPS, 1998; Citeseer: 1998; pp 536-542.

[19] Scholkopf, Bernhard, and Alexander J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond.* MIT press: 2002;98(3): pp 489-489.

[20] Roweis, Sam T., and Lawrence K. Saul. *Nonlinear dimensionality reduction by locally linear embedding*. SCIENCE 2000, 290(5), 2323-2326.

[21] Saul, Lawrence K., and Sam T. Roweis. *Think globally, fit locally: unsupervised learning of low dimensional manifolds*. The Journal of Machine Learning Research 2003, 4(2), 119-155.

[22] Belkin, Mikhail, and Partha Niyogi. *Laplacian eigenmaps for dimensionality reduction and data representation.* NEURAL COMPUT 2003, 15(6), 1373-1396.

[23] Donoho, David L., and Carrie Grimes. *Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data.* Proceedings of the National Academy of Sciences 2003, 100(10), 5591-5596.

[24] Niyogi, X. *Locality preserving projections*, Neural information processing systems, 2004; p 153-161.

[25] Silva, Vin D., and Joshua B. Tenenbaum. *Global versus local methods in nonlinear dimensionality reduction*, Advances in neural information processing systems, 2002; 2002; pp 705-712.

[26] Burges, Christopher JC. *Dimension reduction: A guided tour.* MACH LEARN 2009, 2, 275-365.

PoS(ISC2015)027