

Advanced Control in Micro-/Nanosystems

Guest Editors: Qingsong Xu, Pak-Kin Wong, Minping Jia,
Micky Rakotondrabe, and Li Zhang





Advanced Control in Micro-/Nanosystems

Journal of Control Science and Engineering

Advanced Control in Micro-/Nanosystems

Guest Editors: Qingsong Xu, Pak-Kin Wong, Minping Jia,
Micky Rakotondrabe, and Li Zhang



Copyright © 2012 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in “Journal of Control Science and Engineering.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

Edwin K. P. Chong, USA
Ricardo Dunia, USA
Nicola Elia, USA
Peilin Fu, USA
Bijoy K. Ghosh, USA
F. Gordillo, Spain
Shinji Hara, Japan

Seul Jung, Republic of Korea
Vladimir Kharitonov, Russia
James Lam, Hong Kong
Derong Liu, China
Tomas McKelvey, Sweden
Silviu-Iulian Niculescu, France
Yoshito Ohta, Japan

Yang Shi, Canada
Zoltan Szabo, Hungary
Onur Toker, Turkey
Xiaofan Wang, China
Jianliang Wang, Singapore
Wen Yu, Mexico
Mohamed A. Zribi, Kuwait

Contents

Advanced Control in Micro-/Nanosystems, Qingsong Xu, Pak-Kin Wong, Minping Jia, Micky Rakotondrabe, and Li Zhang
Volume 2012, Article ID 827479, 2 pages

Processing Chip for Thin Film Bulk Acoustic Resonator Mass Sensor, Pengcheng Jin, Shurong Dong, Hao Jin, and Mengjun Wu
Volume 2012, Article ID 923617, 5 pages

Flexible Surface Acoustic Wave Device with AlN Film on Polymer Substrate, Jian Zhou, Shurong Dong, Hao Jin, Bing Feng, and Demiao Wang
Volume 2012, Article ID 610160, 5 pages

Short-Term Prediction of Air Pollution in Macau Using Support Vector Machines, Chi-Man Vong, Weng-Fai Ip, Pak-kin Wong, and Jing-yi Yang
Volume 2012, Article ID 518032, 11 pages

A Memory-Based Hysteresis Model in Piezoelectric Actuators, Guilin Zhang, Chengjin Zhang, and Jason Gu
Volume 2012, Article ID 498590, 7 pages

Model Predictive Engine Air-Ratio Control Using Online Sequential Relevance Vector Machine, Hang-cheong Wong, Pak-kin Wong, and Chi-man Vong
Volume 2012, Article ID 731825, 15 pages

Modelling and Prediction of Particulate Matter, NO_x, and Performance of a Diesel Vehicle Engine under Rare Data Using Relevance Vector Machine, Ka In Wong, Pak Kin Wong, and Chun Shun Cheung
Volume 2012, Article ID 782095, 9 pages

Editorial

Advanced Control in Micro-/Nanosystems

Qingsong Xu,¹ Pak-Kin Wong,¹ Minping Jia,² Micky Rakotondrabe,³ and Li Zhang⁴

¹ Department of Electromechanical Engineering, Faculty of Science and Technology, University of Macau, Avenue Padre Tomás Pereira Taipa, Macau

² School of Mechanical Engineering, Southeast University, Nanjing, Jiangsu 211189, China

³ Automatic Control and Micro-Mechatronic Systems (AS2M) Department, FEMTO-ST Institute, UMR CNRS 6174-UFC/ENSMM/UTBM, 24 rue Alain Savary, 25000 Besancon, France

⁴ Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong

Correspondence should be addressed to Qingsong Xu, qsxu@umac.mo

Received 10 December 2012; Accepted 10 December 2012

Copyright © 2012 Qingsong Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Micro-/nanosystems have attracted plenty of attention at both academia and industry within the last two decades. Micro-/nanosystems include miniaturized systems whose physical dimensions lie in micro-/nanoscale and macro-/mesoscale systems which perform tasks at the micro-/nanoscale. Such systems find extensive applications in micro-/nanomanipulation, micro-/nanoassembly, nanometer-resolution imaging and metrology, health and environmental monitoring, data storage, material science, biomedicine and biotechnology, and so forth. The scale and performance requirement of micro-/nanosystems present a number of challenges to the control system design. In the micro-/nanoworld, the displacements ranging from nanometers to tens of microns and forces from piconewtons to tens of micronewtons are mostly involved. The system performances are very sensitive to external environmental conditions in terms of vibration, temperature, humidity, air velocity, and so on. Due to the precision requirement, the control system design for the micro-/nanoworld concerns with actuation and sensing, monitoring and modeling, dynamics, characterization, position/force measurement techniques, signal processing, controller design, motion planning, hardware implementation, material fabrication, and so on.

The objective of this special issue is to report some most recent developments and contributions in the micro-/nanosystems control. After a thorough review process, a total of 6 out of the submitted papers have been accepted for this special issue and present interesting results.

Piezoelectric actuators are widely used in nano-/micro-positioning systems owing to their attractive properties of fast frequency response, nanometer scale resolution, and high stiffness. In order to achieve a precise positioning, the hysteresis nonlinearity needs to be well understood and suppressed. To tackle this issue, G. Zhang et al. develop a memory-based parabola model to capture the hysteresis behavior using the mathematical transformation. A new converging point was updated to compensate for the prediction error when the hysteresis path hits the upper converging point. The experimental and simulation results demonstrate that the prediction accuracy of the proposed model is noticeably improved as compared with the model without memory operator.

Air pollution is one of the most challenging problems in many cities nowadays. The increased use of motor vehicles causes the amount of exhaust emissions to increase dramatically, which makes the problem more serious. Diesel engines are used extensively in buses and trucks; thus they are the major roadside emitters, posing a significant threat to the health of the road users. In order to reduce these emissions, the combustion process of the engines has to be controlled. A new approach for modeling the performance and emissions of diesel engine is proposed by P.-K. Wong et al. The adopted relevance vector machine (RVM) method is stated clearly and a feasible learning algorithm is given with technical details. The conducted experimental comparative studies demonstrate the superiority of proposed RVM modeling

approach over the conventional ANN ones. This paper represents a new attempt using RVM.

The status of microparticle in air is crucial for our daily life. Forecasting of air pollution is a popular and important topic in recent years due to the health impact caused by air pollution. It is necessary to build an early warning system, which provides forecast and also alerts health alarm to local inhabitants by medical practitioners and the local government. C. M. Vong et al. propose the air pollutant regression modeling and time series prediction using support vector machines technique. The air pollutant modeling and prediction for Macau area can also be extended to other areas. By acquiring the real data, five SVM models with different kernel functions are established and their performances are well investigated. With a series of comparative studies, it is concluded that the SVM with linear or RBF kernel is capable of predicting the air pollutant accurately.

Engine power, brake-specific fuel consumption, and emissions relate closely to air ratio (i.e., lambda) among all the engine variables. An accurate and adaptive model for lambda prediction is essential to effective lambda control for long term. H. C. Wong et al. present a new model predictive control algorithm for air ratio regulation based on an emerging technique, relevance vector machine (RVM). This work gives good supporting evidence that RVM predictive controller may be used as a useful scheme to replace the conventional proportional-integral controller. The methodology has been clearly described while RVM model for engine air-ratio had been successfully implemented and tested on a real car with a satisfied performance, on which RVM model had not been done before.

Flexible electronics is an emerging technology with a tremendous challenge, but has a great potential for applications in display, portable electronics, healthcare, and so forth. So far, flexible electronics has been limited to electronic devices such as diodes and transistors. J. Zhou et al. present the fabrication of high frequency surface acoustic wave (SAW) devices on AlN films, which have potentials in flexible microfluidic applications. The structural properties of AlN films on polymer substrate are also characterized. The obtained resonant frequencies of the SAW devices show no severe deterioration in acoustic speed on the soft substrate, and the results are consistent with the theoretical modeling.

Finally, aiming at portable applications, P. Jin et al. present a new integrated tiny mass sensor based on the thin film bulk acoustic resonator (FBAR) and CMOS technology. The fabricated FBAR has resonance frequency of 1.878 GHz and Q factor of 1200. The precision of the whole processing chip is 1 KHz with the FBAR frequency gap from 25 kHz to 25 MHz. The whole FBAR signal processing circuit is verified with 0.18 μm RF/Mixed-signal CMOS process. It is shown that the size of the entire chip with pads is only 1300 $\mu\text{m} \times 950 \mu\text{m}$, which can be connected with FBAR to process its RF sensor signal and show mass change value directly.

We hope that the readers will find the special issue interesting and stimulating and expect that the included papers contribute to the further advance in the domain of

micro-/nanosystems control. In addition, we would like to express our heartfelt thanks to all the authors who have submitted their papers and all the reviewers who helped handling the papers for this special issue.

Qingsong Xu
Pak-Kin Wong
Minping Jia
Micky Rakotondrabe
Li Zhang

Research Article

Processing Chip for Thin Film Bulk Acoustic Resonator Mass Sensor

Pengcheng Jin, Shurong Dong, Hao Jin, and Mengjun Wu

Department of ISEE, Zhejiang University, Hangzhou 310058, China

Correspondence should be addressed to Shurong Dong, dongshurong@zju.edu.cn

Received 12 January 2012; Revised 22 May 2012; Accepted 8 July 2012

Academic Editor: Li Zhang

Copyright © 2012 Pengcheng Jin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aimed at portable application, a new integrated process chip for thin film bulk acoustic resonator (FBAR) mass sensor is proposed and verified with 0.18 μm CMOS processing in this paper. The longitudinal mode FBAR with back-etched structure is fabricated, which has resonant frequency 1.878 GHz and Q factor 1200. The FBAR oscillator, based on the current-reuse structure, is designed with Modified Butterworth Van Dyke (MBVD) model. The result shows that the FBAR oscillator operates at 1.878 GHz with a phase noise of -107 dBc/Hz and -135 dBc/Hz at 10 KHz and 100 KHz frequency offset, respectively. The whole process chip size with pads is $1300 \mu\text{m} \times 950 \mu\text{m}$. The FBAR and process chip are bonded together to sense tiny mass. The measurement results show that this chip precision is 1 KHz with the FBAR frequency gap from 25 kHz to 25 MHz.

1. Introduction

In recent years, mass sensor based on FBAR technology has a rapid development due to its high mass sensitivity and integrated potential [1]. FBAR mass sensor is considered as an excellent portable healthcare sensor resolution [2, 3]. Recent researches mainly focus on relative humidity sensor [4], glycerol detector [5], gravimetric sensing [6], ultraviolet sensor [7], DNA and protein detection [8], microfluidic system [9], and so on. However, by now FBAR mass sensors are handled with network analyzer and RF probe station, which is massive and various with testing environment. It is not fit to portable application. There is no paper to report FBAR sensor signal processing chip and its FBAR oscillator. In this paper, we presented an integrated chip, which can be connected with FBAR to process its RF sensor signal and show mass change value directly.

2. FBAR Sensor Design

2.1. System Scheme. FBAR structure is shown in Figure 1. It consists of an AlN thin piezoelectric film sandwiched by two Al metal electrodes with back-etched structure and an adsorption layer which is used to adsorb a particular material

for sensor. The adsorption layer should be selected according to the different detected target materials. The resonant frequency will change due to the change of mass of the target material. Conversion between frequency change and the mass loading is described by the Sauerbrey equation [10]:

$$\Delta f_s = -2\Delta m \cdot f_s^2 A^{-1} \mu_q^{-1/2} \rho_q^{-1/2}, \quad (1)$$

in which Δf_s is the frequency change, f_s is the fundamental resonant frequency, Δm is the mass change, A is the active area, ρ_q is the density, and μ_q is the shear modulus.

The sensor process system should be designed to obtain FBAR resonant frequency changes due to tiny mass. Usually, the system is based on dual-path structure, shown in Figure 2, one path for sensor signal and the other for reference signal to deembed testing environment effects change, such as pressure, temperature, and humidity, due to the two paths that are neighbor in one chip. Because the signals of FBAR are always weak, an oscillator network is designed to active FBAR signals to obtain 3V output to drive the following processing circuit. The FBAR resonant frequency is usual about 2-3 GHz. It is too high to be processed with normal high speed counter. Mixture frequency approach is also not fit to FBAR sensor owing it IP3 and complex

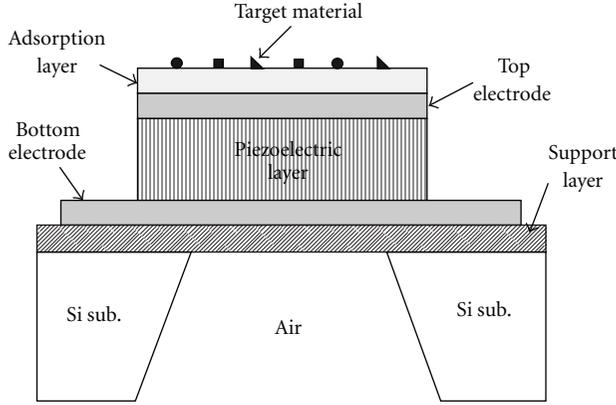


FIGURE 1: Cross-section structure of FBAR.

structure. The designed new chip employs two divide-by-256 dividers to lower FBAR resonant frequencies to about 10 MHz, then it counted them by two high speed counters. The frequency difference is obtained by the subtractor. According to this counting difference, the change of mass load can be obtained from the following calculation:

$$\Delta m \approx -\frac{M}{f_s} \cdot \Delta f = -\frac{M \cdot N}{f_s \cdot T} \cdot S, \quad (2)$$

where M is the quality of the work area of FBAR piezoelectric, N is the division ratio of divider, T is the counting cycle of counter, and S is the frequency counting difference. The whole chip schematic also includes self-calibration and UART interface.

2.2. FBAR and Its Oscillator. Back-etched structure is used to manufacture FBAR, because it is relatively easy to fabricate and has better performance compared to solid-mounted structure (SMR). Figure 3(a) is the top view of the FBAR, and the measured S parameter is shown in Figure 3(b). It shows that the measured Q factor of FBAR is above 1200, and the resonant frequency is 1.878 GHz. The FBAR oscillator is composed of a FBAR, a pair of NMOS and PMOS, and some resistors and capacitors to form the current-reuse structure, in which the negative conductance is supplied by the pair of MOS transistors, as shown in Figure 4. Compared to the traditional cross-couple structure, this configuration consumes less power since the pair of NMOS and PMOS switches simultaneously, while the pair of NMOSs or PMOSs in the conventional structure converts alternately. FBAR oscillators designed and fabricated with 0.18 μm RF/mixed-signal CMOS process and the MBVD model are employed [11]. The simulation result of phase noise of the FBAR oscillator is shown in Figure 5, and the result shows that the oscillator operates at 1.878 GHz with a phase noise of -107 dBc/Hz and -135 dBc/Hz at 10 KHz and 100 KHz frequency offset, respectively.

2.3. Sensor Signal Processing Circuit. According to the operation of FBAR oscillators, we get the output frequency between 1.5 GHz and 2.0 GHz. In order to meet the digital

signal processing requirements, two divide-by-256 dividers are designed to reduce the frequencies down to 10 MHz. Owing to the high frequency, the front-end of dividers should adopt high speed circuits. Hence, signal-coupled logic (SCL) structure which is the evolution of emitter-coupled logic (ECL) is used due to its high speed, low power, and low noise [12]. The divide-by-256 divider consists of a single-ended to differential converter, three cascaded divide-by-2 SCL dividers, and five cascaded divide-by-2 D-flip-flop (DFF) dividers. The reason why it does not implement with all SCL dividers is that the layout area of SCL is larger than DFF structure. The main structure of SCL divider is master-slaver D flip-flop consisting of two D-latches. Figures 6(a) and 6(b) show the block diagram of the divide-by-2 SCL divider and the circuit schematic of the D-latch, respectively. M1 and M4 constitute the sampling circuit, and the cross-coupled pair M2 and M3 constitutes the holding circuit. When the clock signal (CK) is at high state, D-Latch works like a buffer; when the clock signal is low, the cross-coupled pair holds the existing state through positive feedback principle. Additionally, five stages of divide-by-2 DFF divider are implemented based on the traditional master-slaver structure.

The signal after divide-by-256 will be sent into standard digital signal processing circuit. In this paper, 40 MHz temperature compensate X'tal (crystal) oscillator (TCXO) is used as the high precision clock due to its high frequency stability, wide frequency range, and high frequency accuracy. Taking into account the power consumption, the sampling period of this system is set to 500 ms, and the counting period is set to 256 ms. The word length of the count difference is 24 bits. Figure 7 gives the top block diagram of the digital processing circuit generated by Verdi. It consists of timing module, counter module, subtractor module, encoding module, UART clock module, and sending module.

3. Measurement Results

The whole FBAR signal processing circuit is verified in 0.18 μm RF/mixed-signal CMOS process, and the dimension of the chip with pads is 1300 $\mu\text{m} \times 950 \mu\text{m}$. Figure 8 shows the chip micrograph. The chip is banded with FBAR by gold wire, and the whole sensor system is tested. The measurement results are listed in Table 1, where the frequency difference ranges from 25 kHz–25 MHz. This tiny mass sensor signal processing circuit has high accuracy, but sometimes there is an accidental error of ± 1000 Hz, which is related to the synchronization of high speed counter trigger. For example, when the signal frequency difference is 100 kHz, the measured difference is 100 kHz in most case, but sometimes it is 99 kHz or 101 kHz.

4. Conclusion

An integrated tiny mass sensor based on FBAR and CMOS technology is proposed and verified in this paper. The system of FBAR sensor is designed with dual-path structure. One path is used for sensor signal, and the other for reference signal to diminish the effect of environment change. FBAR

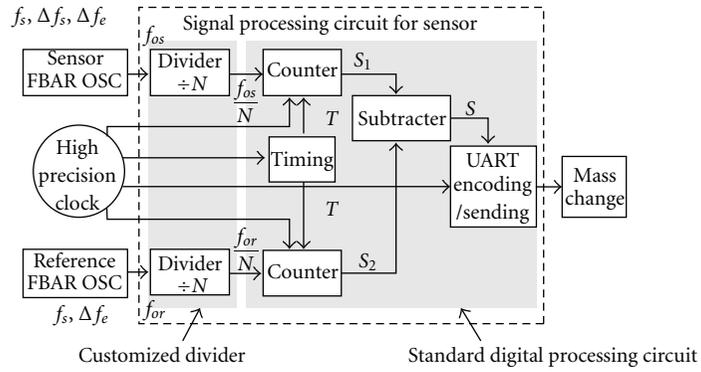


FIGURE 2: Scheme of FBAR tiny sensor.

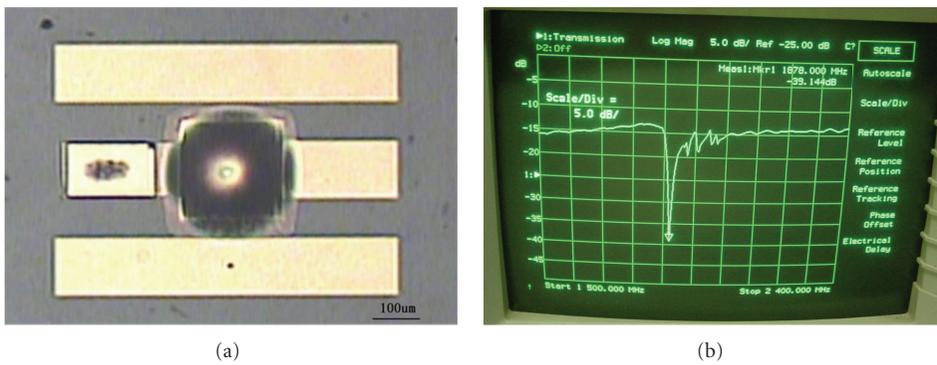


FIGURE 3: (a) Chip photo of the fabricated FBAR, (b) measured resonant frequency of FBAR.

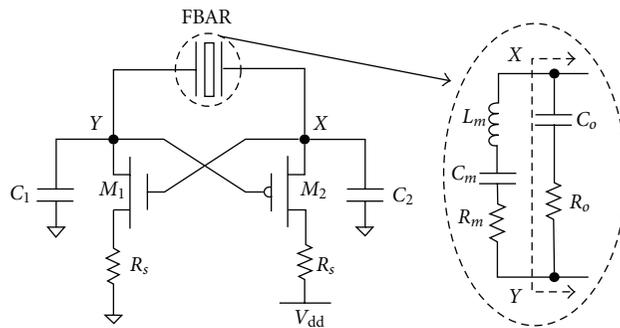


FIGURE 4: Schematic of the FBAR oscillator and equivalent circuit of FBAR.

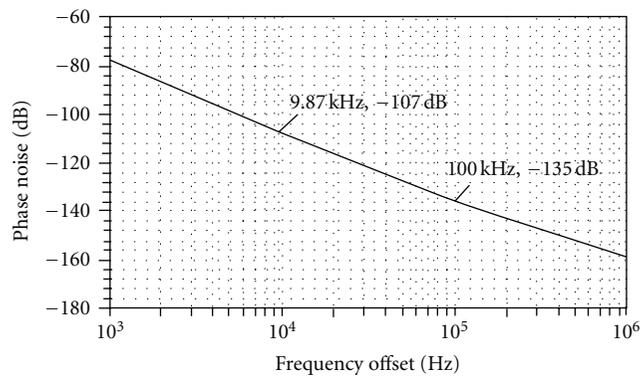


FIGURE 5: Phase noise of the FBAR oscillator from postlayout simulation.

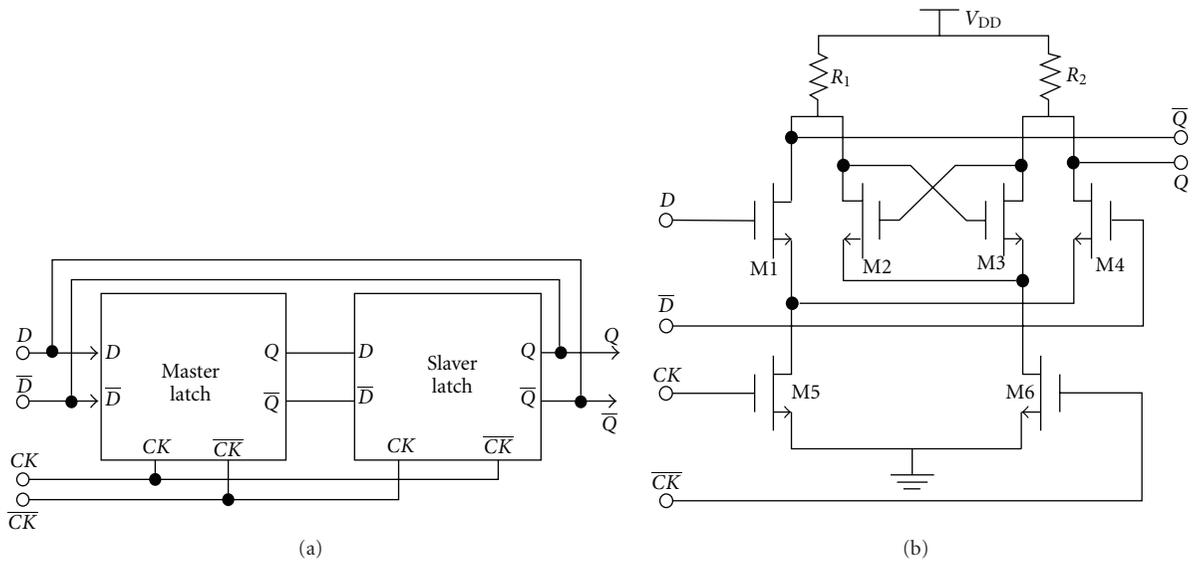


FIGURE 6: (a) Block diagram of divide-by-2 SCL divider, (b) circuit schematic of D-latch.

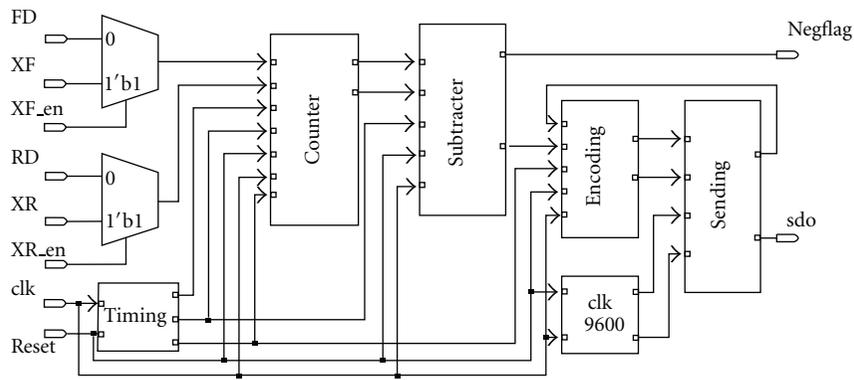


FIGURE 7: Block diagram of the digital processing circuit generated by Verdi.

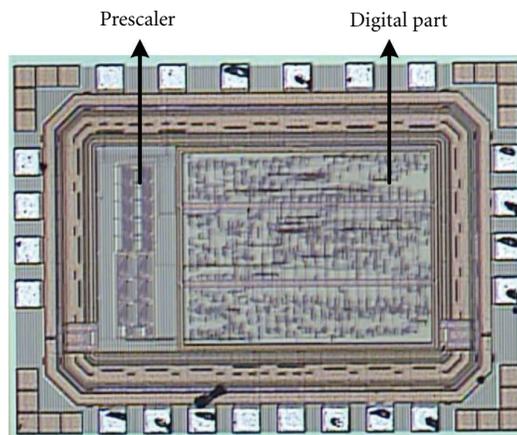


FIGURE 8: Chip micrograph.

TABLE 1: Measurement results of sensor signal processing circuit.

Reference signal/Hz	Sensor signal/Hz	Measured difference/Hz	Accidental error/Hz
1 899 000 000	1 898 975 000	25 000	±1 000
1 899 000 000	1 898 900 000	100 000	±1 000
1 899 000 000	1 889 000 000	10 000 000	±1 000
1 899 000 000	1 874 000 000	25 000 000	±1 000

operates in longitudinal mode, and its resonant frequency is 1.878 GHz with Q factor above 1200. Two FBAR signals are activated by oscillators based on current-reuse differential configuration to promote their output signals. Subsequently, these two FBAR oscillator signals are divided by 256 and then sent to digital signal processing circuit to obtain the frequency difference. Finally, this frequency offset is used to evaluate the tiny mass loading change. The whole FBAR signal processing circuit is verified in 0.18 μm RF/mixed-signal CMOS process.

Acknowledgment

This work was supported by a grant from the National Natural Science Foundation Key Program of China (no. 61171038, 60936002).

References

- [1] R. Ruby, "Review and comparison of bulk acoustic wave FBAR, SMR technology," in *Proceedings of the IEEE Ultrasonics Symposium (IUS'07)*, pp. 1029–1040, October 2007.
- [2] G. Wingqvist, "AlN-based sputter-deposited shear mode thin film bulk acoustic resonator (FBAR) for biosensor applications—a review," *Surface and Coatings Technology*, vol. 205, no. 5, pp. 1279–1286, 2007.
- [3] R.-C. Lin, Y.-C. Chen, W.-T. Chang, C.-C. Cheng, and K.-S. Kao, "Highly sensitive mass sensor using film bulk acoustic resonator," *Sensors and Actuators A*, vol. 147, no. 2, pp. 425–429, 2008.
- [4] X. T. Qiu, R. Tang, J. Zhu et al., "Experiment and theoretical analysis of relative humidity sensor based on film bulk acoustic-wave resonator," *Sensors and Actuators B*, vol. 147, no. 2, pp. 381–384, 2010.
- [5] M. Link, J. Weber, M. Schreiter, W. Wersing, O. Elmazria, and P. Alnot, "Sensing characteristics of high-frequency shear mode resonators in glycerol solutions," *Sensors and Actuators B*, vol. 121, no. 2, pp. 372–378, 2007.
- [6] S. Rey-Mermet, R. Lanz, and P. Muralt, "Bulk acoustic wave resonator operating at 8 GHz for gravimetric sensing of organic films," *Sensors and Actuators B*, vol. 114, no. 2, pp. 681–686, 2006.
- [7] X. Qiu, J. Zhu, J. Oiler, C. Yu, Z. Wang, and H. Yu, "Film bulk acoustic-wave resonator based ultraviolet sensor," *Applied Physics Letters*, vol. 94, no. 15, Article ID 151917, 2009.
- [8] M. Nirschl, A. Blüher, C. Erler et al., "Film bulk acoustic resonators for DNA and protein detection and investigation of in vitro bacterial S-layer formation," *Sensors and Actuators A*, vol. 156, no. 1, pp. 180–184, 2009.
- [9] G. Sharma, L. Liljeholm, J. Enlund, J. Bjurström, I. Katardjiev, and K. Hjort, "Fabrication and characterization of a shear mode AlN solidly mounted resonator-silicone microfluidic system for in-liquid sensor applications," *Sensors and Actuators A*, vol. 159, no. 1, pp. 111–116, 2010.
- [10] D. S. Ballantine, R. M. White, S. J. Martin et al., *Acoustic Wave Sensors: Theory, Design, and Physico-Chemical Application*, Academic Press, 1997.
- [11] W. W. Cheng, S. R. Dong, Y. Han et al., "A low power, low phase noise fbar oscillator," *Integrated Ferroelectrics*, vol. 105, no. 1, pp. 75–86, 2009.
- [12] N. Krishnapura and P. R. Kinget, "A 5.3-GHz programmable divider for HiPerLAN in 0.25- μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 7, pp. 1019–1024, 2000.

Research Article

Flexible Surface Acoustic Wave Device with AlN Film on Polymer Substrate

Jian Zhou, Shurong Dong, Hao Jin, Bing Feng, and Demiao Wang

Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China

Correspondence should be addressed to Shurong Dong, dongshurong@zju.edu.cn

Received 12 January 2012; Revised 25 May 2012; Accepted 8 July 2012

Academic Editor: Li Zhang

Copyright © 2012 Jian Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Surface acoustic wave device with *c*-axis-oriented aluminum nitride (AlN) piezoelectric thin films on polymer substrates can be potentially used for development of flexible sensors, flexible microfluidic applications, microsystems, and lab-on-chip systems. In this work, the AlN films have been successfully deposited on polymer substrates using the DC reactive magnetron-sputtering method at room temperature, and the XRD, SEM, and AFM methods reveal that low deposition pressure is beneficial to the highly *c*-axis-oriented AlN film on polymer substrates. Studies toward the development of AlN thin film-based flexible surface acoustic wave devices on the polymer substrates are initiated and the experimental and simulated results demonstrate the devices showing the acoustic wave velocity of 9000–10000 m/s, which indicate the AlN lamb wave.

1. Introduction

The surface-acoustic-wave- (SAW-) based microfluidic devices can be used not only for pumping, mixing, and droplet generation but also for biosensors and single-mechanism-based lab-on-a-chip applications [1]. By now, all SAW devices are fabricated on the stiff substrates instead of flexible's, such as LiNbO₃ [2], Piezoelectric (PE) thin films on Si wafer [3, 4], diamond [5], Al₂O₃ [6] and so on. SAW on flexible substrate is cheaper and can be bent easily, which is fitted for portable microfluidic applications, such as wrist health-care monitor. In this paper, a SAW device with aluminum nitride (AlN) piezoelectric thin films is fabricated on the polymer substrates, and its resonance response is also investigated.

2. Experimental Details

Kapton polyimide film 100H (Dupont-Toray Inc., thickness 250 μm) was chosen as the flexible substrate owing to its excellent mechanical and electrical properties, chemical stability, and wide operating temperature range (−269°C to +400°C). AlN was deposited by a home-made DC

magnetron-sputtering system. The base pressure of the chamber was 1×10^{-4} Pa before deposition. The aluminum (Al) target of purity was 99.999% and a diameter of 20 mm was used and water-cooled. The distance between the target and the substrate was fixed at 70 mm. Before depositing the AlN film, an Al underlayer was deposited on the polyimide substrate as a transition with deposition pressure of 0.27 Pa and DC power of 300 W. AlN was then deposited on the Al-coated polyimide substrate in an N₂/Ar atmosphere. The effect of deposition pressure on the properties of the AlN films was investigated. The AlN deposition time for all samples is one hour, and the substrates were not intentionally heated.

The crystalline structure and crystal orientation of films were analyzed by X-ray diffraction (XRD-6000, Japan) using Cu-K_α radiation and scanned angle of $2\theta = 20^\circ \sim 70^\circ$. The degree of *c*-axis crystallization was examined by the full width at half maximum (FWHM) of the AlN (002) diffraction peak. The strain and crystallite size of the thin film were extracted from the XRD data by the standard method. Strain is calculated from $\epsilon_z = (c - c_0)/c_0$ [7], where c_0 is the strain-free lattice parameter (4.979 Å) and the lattice constant c is equal to twice the interplanar spacing d ,

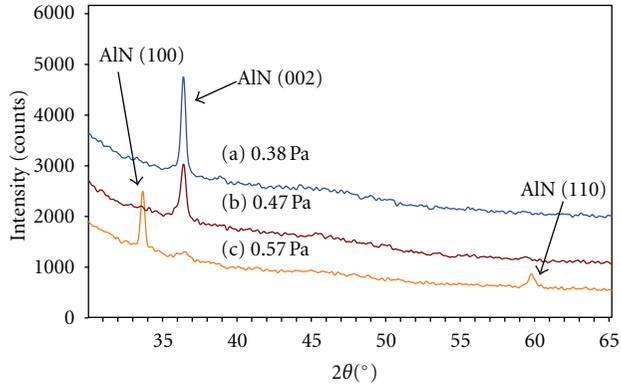


FIGURE 1: XRD patterns of AlN thin films deposited on polymer substrates with Al electrodes under different deposition pressures.

measured from the position of the (002) peak using Bragg-equation. Crystallite sizes were calculated from the Debye-Scherrer formula [8]: $D = K\lambda/(\beta \cos \theta)$, where K is the shape factor of the average crystallite with value of 0.94, λ the X-ray wavelength (0.15406 nm for Cu target), β the FWHM in radians, θ the Bragg angle, and D the mean crystallite dimension normal to diffracting planes. For cross-sectional columnar structure observation, the microscopic film was observed using a Scanning Electron Microscope (SEM) (S4800, Hitachi, Ltd., Japan). The surface morphology and root mean square (rms) surface roughness of the AlN films were measured by atomic force microscopy (AFM) (SPI-3800N, Seiko, Japan).

To study the SAW propagation characteristics on flexible substrates, two-port resonators were designed and fabricated by conventional deep UV photolithography and lift-off process. Al was used as the electrodes with a thickness of 150 nm and each transducer consisted of 10 pairs of IDTs. The distance between the two transducers was 10λ , where the SAW wavelength λ is determined by the IDT pitch. The aperture was 80λ and the distance between the IDTs and the adjacent shorted reflecting gratings was designed as 5.375λ to ensure the formation of a standing wave at center frequency. The frequency characterization was carried out using Agilent 8722ES network analyzer.

3. Results and Discussion

Figure 1 shows the XRD pattern of the AlN films deposited at different deposition pressures. In Figures 1(a) and 1(b), AlN film shows a main XRD peak near $2\theta = 36.1^\circ$ which corresponds to the AlN (002) crystal orientation. The results demonstrate that the AlN crystal structures are perpendicular to the polymer substrate with a good (002) orientation. The FWHM of the AlN (002) peak with 0.38 Pa is 0.321° . The grain size is 274 \AA and the strain is 0.4%. Figure 2 is an SEM micrograph of the cross-sectional structure of the AlN films deposited on the polymer substrate at the deposition pressure of 0.38 Pa. It is obvious that the film shows a neat arrangement and exhibits a typical (002)-oriented columnar structure. The surface morphologies and

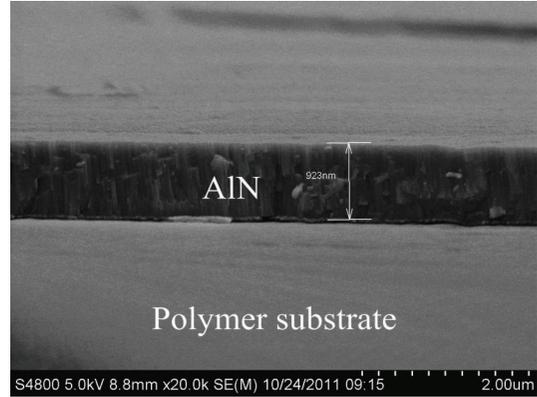


FIGURE 2: The SEM micrograph of AlN thin films on polymer substrate.

the rms surface roughness of the AlN are measured by AFM, as shown in Figure 3. With the increase of the deposition pressure, the rms surface roughness increases. The smoothest AlN film is obtained at the deposition pressure of 0.38 Pa with the rms surface roughness of 4.8 nm.

Two-port resonators (Figure 4) have been fabricated on the AlN film deposited with an N_2/Ar flow ratio of 1 : 2 and pressure of 0.38 Pa. The AlN film for the resonators has the thickness of $1.23 \mu\text{m}$, FWHM of 0.321° and grain size of 274 \AA . Two types of SAW samples were fabricated with different wavelengths of $7.128 \mu\text{m}$ and $6 \mu\text{m}$, respectively. The resonance frequency of a SAW device is determined by the equation $f_0 = V_p/\lambda$, where f_0 is the central frequency, V_p the phase velocity of the acoustic wave, and λ the acoustic wavelength.

Figures 5(a) and 5(b) present the measured S_{21} spectra of the fabricated SAW resonators with the wavelengths of $7.128 \mu\text{m}$ and $6 \mu\text{m}$, respectively. The measured S_{21} spectra take the form of the Sinc function as expected for uniform IDT SAW devices [9]. The resonance frequency of the SAW resonator with the wavelength of $7.128 \mu\text{m}$ is 1.355 GHz, corresponding to the acoustic wave velocity of 9658 m/s, whereas the resonance frequency of the SAW resonator with the wavelength of $6 \mu\text{m}$ shifts to 1.605 GHz, corresponding to the acoustic wave velocity of 9630 m/s, showing that the acoustic wave velocity is not affected by the device structure. The results have clearly demonstrated the piezoelectric effect of AlN film on polymer substrate.

To confirm whether the resonant frequency is generated through PE effect, the frequency response of the resonators has been modeled by the commercial software COMSOL Multiphysics. As the simulation diameters, the thicknesses of Al IDTs, AlN film, Al underlayer, and polyimide substrate are set to be the same as used in devices and are 150 nm, $1.23 \mu\text{m}$, 70 nm, and 100 μm , respectively. Figures 6(a) and 6(b) show the simulated results of the SAW resonators with the wavelengths of $7.128 \mu\text{m}$ and $6 \mu\text{m}$, respectively. Since the resonance frequency has the largest total stored energy, the SAW resonator with the wavelength of $7.128 \mu\text{m}$ has the resonance frequency of 1.415 GHz, corresponding to the acoustic wave velocity of 10086 m/s, and that with

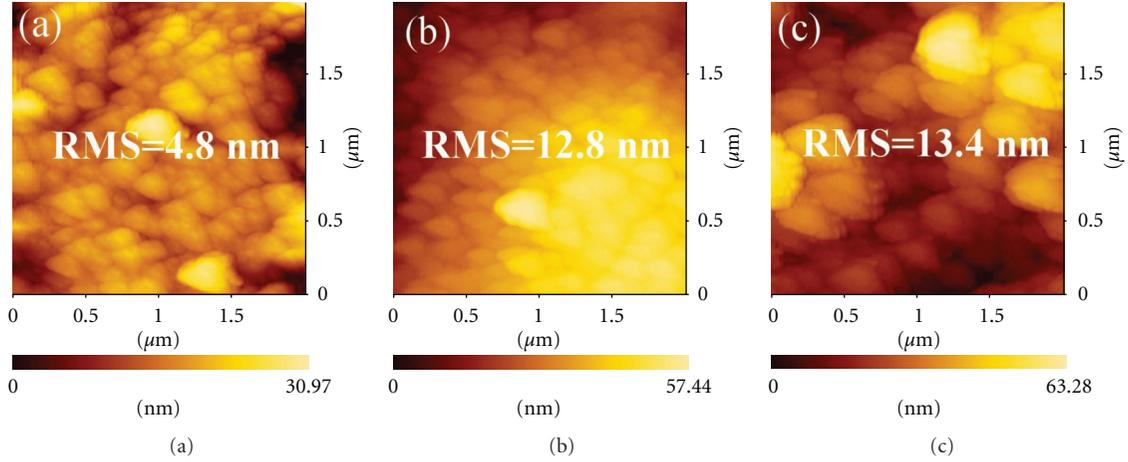


FIGURE 3: Surface morphologies of AlN films on polymer substrates under different deposition pressures: (a) 0.38 Pa, (b) 0.47 Pa, and (c) 0.57 Pa.

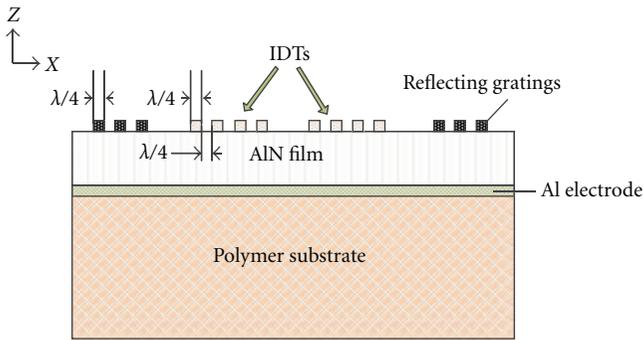


FIGURE 4: The structure of 2-port SAW resonator on polymer substrate.

the wavelength of $6\mu\text{m}$ has the resonance frequency of 1.66 GHz, corresponding to the acoustic wave velocity of 9960 m/s. The experimental results are in good agreement with the simulated results.

Polymer substrates have a very low acoustic impedance (2 Mrayls) which is much smaller than that of the AlN layer (36 Mrayls). The Al underlay enhances the electromechanical coupling coefficient as an electric field can be induced between the IDT electrodes and the Al underlay [10]. Moreover, the c -axis-oriented thin AlN films have relatively large thicknesses at about $d = 0.2\lambda$, where λ is the acoustic wavelength; therefore, the detected acoustic waves are the symmetrical S_0 Lamb waves which theoretically have the acoustic wave velocity near 10,000 m/s [11–13] in AlN. The velocity of the experimental results is slightly smaller than the simulated and theoretical results, possibly due to the defects of the AlN film and the slow-down effects by the Al-coated polyimide substrates. Figures 7(a) and 7(b) show the simulated results of AlN without Al and polyimide substrates. It is clear that the resonance frequency of the SAW resonator with the wavelength of $7.128\mu\text{m}$ is 1.524 GHz, corresponding to the acoustic wave velocity of 10863 m/s, whereas the resonance frequency of the SAW resonator with

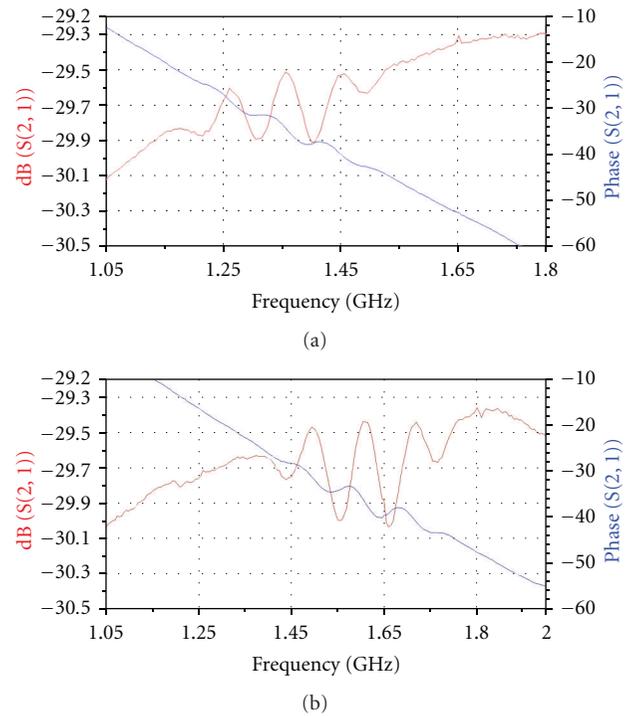


FIGURE 5: The measured S_{21} spectrum of the fabricated SAW resonators with a wavelength of (a) $7.128\mu\text{m}$ and (b) $6\mu\text{m}$.

the wavelength of $6\mu\text{m}$ shifts to 1.81 GHz, corresponding to the acoustic wave velocity of 10860 m/s; both are higher than those in SAW with Al-coated polymer substrate. This result shows that Al-coated polyimide substrates would decrease the resonance frequency due to the low wave velocity of Al-coated polyimide substrate.

4. Conclusions

In this research, we have synthesized and characterized the AlN thin films on Dupont Kapton polyimide substrates by

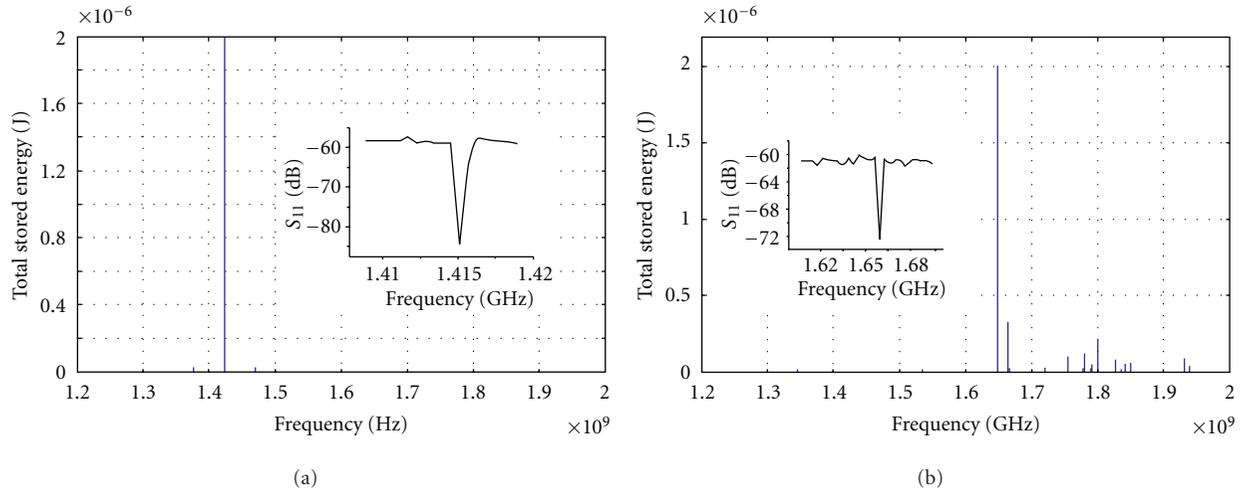


FIGURE 6: The simulated resonance response of the SAW resonators with the wavelengths of (a) $7.128 \mu\text{m}$ and (b) $6 \mu\text{m}$.

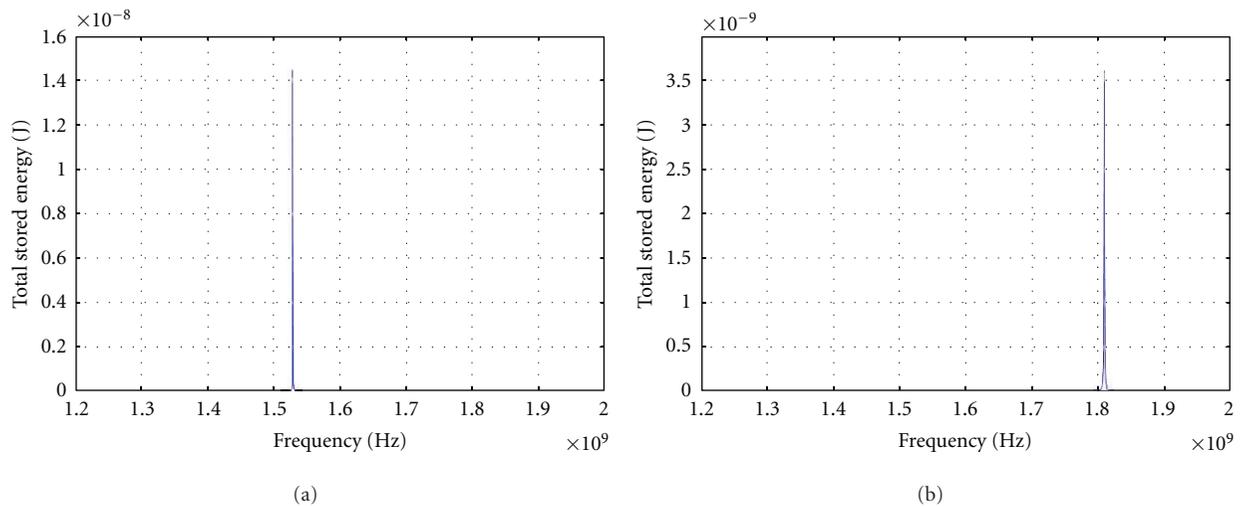


FIGURE 7: The simulated resonance response of the SAW resonators without Al and polyimide substrates with the wavelengths of (a) $7.128 \mu\text{m}$ and (b) $6 \mu\text{m}$.

DC magnetron reactive sputtering. The XRD, SEM, and AFM are used to characterize the orientation, cross-sectional structure, surface morphology, and thickness of AlN thin films. The results show that low deposition pressure is beneficial for AlN (002) orientation. Flexible substrate-based Lamb wave resonators have been fabricated. The experimental and simulated results demonstrate of the devices have the acoustic wave velocity of 9000–10000 m/s, and the deposited AlN film on polyimide substrate are high quality, and have successfully shown the piezoelectric effect.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 61171038), the Research Fund of International Young Scientists (no. 61150110485), and the

Zhejiang Provincial Natural Science Foundation of China (no. Q12F010048).

References

- [1] J. K. Luo, Y. Q. Fu, Y. Li et al., “Moving-part-free microfluidic systems for lab-on-a-chip,” *Journal of Micromechanics and Microengineering*, vol. 19, no. 5, Article ID 054001, 2009.
- [2] C. C. Cheng, K. S. Kao, and Y. C. Chen, “Highly c-axis oriented AlN films deposited on LiNbO₃ substrates for surface acoustic wave devices,” in *Proceedings of the 12th IEEE International Symposium on Applications of Ferroelectrics*, vol. 4, pp. 439–442, August 2000.
- [3] M. Clement, L. Vergara, E. Iborra, A. Sanz-Hervás, J. Olivares, and J. Sangrador, “AlN-on-Si SAW filters: influence of film thickness, IDT geometry and substrate conductivity,” in *Proceedings of the IEEE Ultrasonics Symposium*, vol. 4, pp. 1900–1904, September 2005.

- [4] M. Clement, L. Vergara, J. Sangrador, E. Iborra, and A. Sanz-Hervás, "SAW characteristics of AlN films sputtered on silicon substrates," *Ultrasonics*, vol. 42, pp. 403–407, 2004.
- [5] S. Fujii, "High-frequency surface acoustic wave filter based on diamond thin film," *Physica Status Solidi (A) Applications and Materials*, vol. 208, no. 5, pp. 1072–1077, 2011.
- [6] G. Bruckner, J. Bardong, R. Fachberger, E. Forsén, and D. Eisele, "Investigations of SAW delay lines on c-plane AlN/sapphire at elevated temperatures," in *Proceedings of the IEEE International Frequency Control Symposium (FCS '10)*, pp. 499–502, June 2010.
- [7] W. T. Lim, B. K. Son, D. H. Kang, and C. H. Lee, "Structural properties of AlN films grown on Si, Ru/Si and ZnO/Si substrates," *Thin Solid Films*, vol. 382, no. 1-2, pp. 56–60, 2001.
- [8] B. D. Cullity and S. Stock, *Elements of X-Ray Diffraction*, Prentice Hall, Upper Saddle River, NJ, USA, 3rd edition, 2001.
- [9] C. Campbell, *Surface Acoustic Wave Devices for Mobile and Wireless Communications*, Academic Press, New York, NY, USA, 1998.
- [10] G. Wingqvist, L. Arapan, V. Yantchev, and I. Katardjiev, "A micromachined thermally compensated thin film Lamb wave resonator for frequency control and sensing applications," *Journal of Micromechanics and Microengineering*, vol. 19, p. 035018, 2009.
- [11] J. Bjurström, I. Katardjiev, and V. Yantchev, "Lateral-field-excited thin-film Lamb wave resonator," *Applied Physics Letters*, vol. 86, no. 15, Article ID 154103, pp. 1–3, 2005.
- [12] V. Yantchev and I. Katardjiev, "Design and fabrication of thin film lamb wave resonators utilizing longitudinal wave and interdigital transducers," in *Proceedings of the IEEE Ultrasonics Symposium*, vol. 3, pp. 1580–1583, September 2005.
- [13] J. Bjurström, V. Yantchev, and I. Katardjiev, "Thin film Lamb wave resonant structures-the first approach," *Solid-State Electronics*, vol. 50, no. 3, pp. 322–326, 2006.

Research Article

Short-Term Prediction of Air Pollution in Macau Using Support Vector Machines

Chi-Man Vong,¹ Weng-Fai Ip,² Pak-kin Wong,³ and Jing-yi Yang¹

¹Department of Computer and Information Science, University of Macau, Macau

²Faculty of Science and Technology, University of Macau, Macau

³Department of Electromechanical Engineering, University of Macau, Macau

Correspondence should be addressed to Chi-Man Vong, cmvong@umac.mo

Received 16 January 2012; Accepted 29 March 2012

Academic Editor: Qingsong Xu

Copyright © 2012 Chi-Man Vong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Forecasting of air pollution is a popular and important topic in recent years due to the health impact caused by air pollution. It is necessary to build an early warning system, which provides forecast and also alerts health alarm to local inhabitants by medical practitioners and the local government. Meteorological and pollutions data collected daily at monitoring stations of Macau can be used in this study to build a forecasting system. Support vector machines (SVMs), a novel type of machine learning technique based on statistical learning theory, can be used for regression and time series prediction. SVM is capable of good generalization while the performance of the SVM model is often hinged on the appropriate choice of the kernel.

1. Introduction

Air pollution is often the result of economy development and population increase. It is particular in developing cities, particularly cities in China and India. Many epidemiologic studies [1–3] reported that air pollution problems are often associated with adverse human respiratory health effects, particularly to susceptible individuals. For example, ozone has been attributed to cause inflammation in airway and elevate airway response to inhaled allergens. It may increase the risk of the development of asthma among children taking part in outdoor sports [4]. WHO [5] reported that the health problems in turn may increase the burden of the health care systems in the long run and be detrimental to economy. To reduce the burden on health care due to diseases caused by atmospheric pollutants, the establishment of an early warning system is necessary. The success of an early warning system, which provides forecast and alerts local inhabitants, depends on the reliability and the availability of up-to-date meteorological information and pollutions data. For instance, medical practitioners can advise patients to minimize outdoor activities during days of high levels of pollutions and smog, depending on the prediction of the early warning system.

The meteorological and pollutant data in Macau are used as a case study for the testing of the forecasting model for a representative developing city. Macau, located on the southern coast of China with merely 26.8 square miles land area, comprises three land zones: Macau peninsula, Taipa, and Coloane (Figure 1). Macau peninsula has the characteristics of a hybridized, urbanized area; Taipa has mainly residential areas; Coloane has a power station and is largely undeveloped with the largest green areas. The population density in 2008 reached 20,493 people per square mile [6], which is one of the highest in the world. Resident population in Macau is projected to increase at an average annual rate of 1.9%, from 513,000 at 2006 to 829,000 at 2031 [7]. At present, the number of vehicles amounts to 188, 668 by the end of 2009, that is, more than triple from the quantity in 1999 [8]. Macau Government has implemented regulations on importing lead-free petroleum products inline with other developed countries since 1995, while the sulfur contents in gasoline must be lower than 0.05% by weight. Furthermore, in 2004, the sulfur contents in petroleum products used in power station were also regulated. Implementation of these policies reduces the emission of pollutants locally. In addition, construction and infrastructure projects have been transforming the landscape

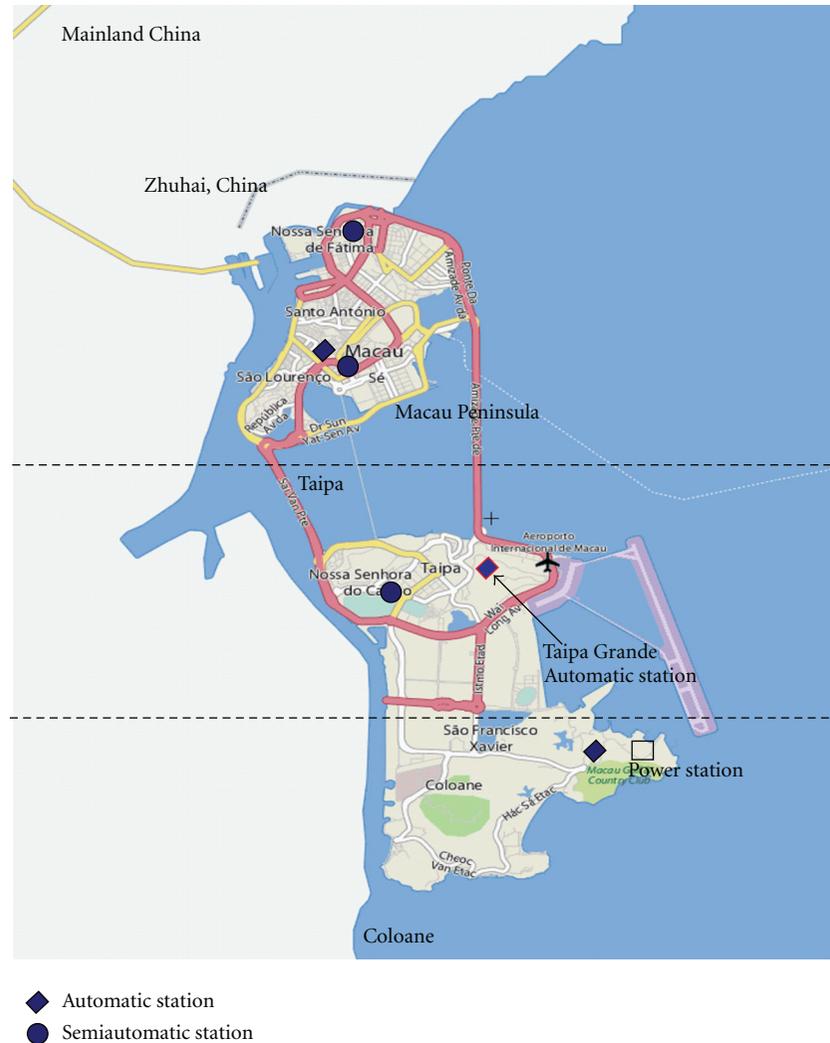


FIGURE 1: The three land zones of Macau and the location of the stations.

of a reclaimed land area between Taipa and Coloane since 2004.

Monitoring and forecasting of air pollutant level in ambient condition involve using a variety of approaches, for example, on-site measurement, computational fluid dynamics (CFD) simulation, and computational intelligence, and so forth. Artificial neural network (ANN) method is regarded as a cost-effective method and has been employed for the construction of prediction models at a variety of cities by environmental researchers [9–12]. The practical applications of these models, however, suffer from different drawbacks, for example, local minima, overfitting, poor generalization, and the need to determine the appropriate network architecture. Support vector machines (SVM), developed by Vapnik [13], can provide an effective novel approach to improve generalization performance of neural networks and achieve global solutions simultaneously. SVM can overcome most drawbacks of ANN and has been reported to show promising results [14–16]. However, the performance of the resulting SVM is often hinged on the appropriate choice of the

kernel. There are several kernels commonly used in SVM for regression. Therefore, another aim of this study is to study which kernel is more suitable used in air pollution prediction.

2. Meteorological Data

The meteorological information and pollutant data measured at Taipa Grande automatic meteorological station (see Figure 1) in year 2003 to year 2006 were selected as the experiment data set, which were extracted from Macau Government's centre. Since the land area of Macau is relatively small, the data obtained at Taipa Grande (at an elevation of approximately 150 m above sea level) may be considered as representative for the entire region of Macau. The meteorological stations record pollutant data, such as nitrogen dioxide (NO_2), sulfur dioxide (SO_2), suspended particulate matters (SPM), and ozone (O_3); climatic data, such as temperature, humidity, rainfall, wind direction, wind

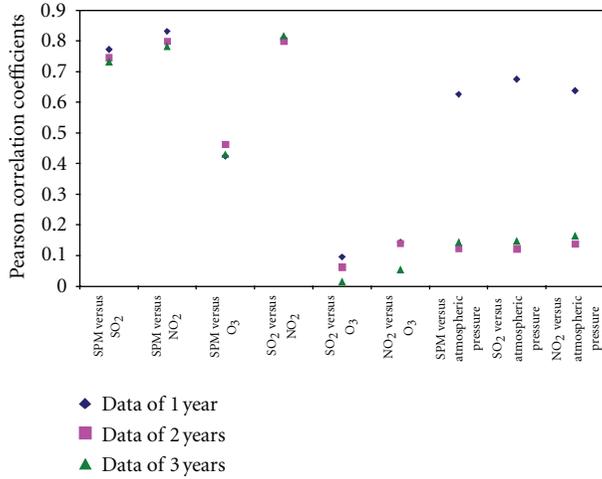


FIGURE 2: Pearson correlation coefficients between the various climatic and pollutant parameters.

TABLE 1: Pearson correlation coefficients between selected parameters.

Corr(X, Y)	Pollutant			
	SPM	SO ₂	NO ₂	O ₃
SO ₂	0.731	—	—	—
NO ₂	0.782	0.817	—	—
O ₃	0.429	0.014	-0.054	—
Temp	-0.387	-0.585	-0.587	0.178
Mean RH	-0.625	-0.528	-0.482	-0.407
Wind speed	-0.096	0.069	-0.083	-0.140
Sunshine Hr	0.081	-0.032	-0.080	0.315*
Air pressure	0.143	0.147	0.164	0.055

Bolded values were used in this model; *sunshine hour has a close relationship with O₃ level, so sunshine hour was selected to be one of the input variables for O₃ model.

speed, and precipitation. The day average value for air pollutants and meteorological data is considered a more representative measure and is adopted in this study.

In addition, the recorded levels of SPM, SO₂, NO₂, and O₃ in January and July 2006 were selected as special cases in this study. The reason to choose the data in these two months is because January and July represent the winter and summer seasons in Macau, respectively. January is typified with dry, dominating northeastern wind, whereas July is typified with humid, hot weather, and southeastern prevailing wind. The temperature difference in these two seasons in Macau may range from 10 to 30°C. In winter season, due to the dominating northeastern wind, air-borne industrial pollutants from mainland China may be blown through Macau. In contrast, the southeastern prevailing wind from the sea in summer usually carries pollutants away.

Moreover, these meteorological data are closely associated with the presence and dispersion of pollutants. In order to discern the relationship between meteorological data and pollutants, an unadjusted crude method of bivariate analysis

TABLE 2: Pearson correlation coefficients between pollutants in different time series.

	Pollutant			
	SPM	SO ₂	NO ₂	O ₃
Lag 0	1.00	1.00	1.00	1.00
Lag 1	0.76	0.68	0.76	0.78
Lag 2	0.59	0.53	0.64	0.60
Lag 3	0.51	0.47	0.61	0.51
Lag 4	0.46	0.42	0.57	0.45
Lag 5	0.47	0.41	0.55	0.40

using Pearson correlation coefficients was applied. The resultant Pearson correlation coefficients between the various meteorological and pollutant parameters are presented in Figure 2. The results illustrate some critical problems using limiting amount of data for the determination of the relationship between bivariate independent variables. For instance, atmospheric pressure appeared to have a positive correlation with SPM, NO₂, and SO₂ for a period of one year, respectively, whereas the same parameters for a three-year period showed no relationship. However, to minimize the operation of the regression model, parameters with a Pearson correlation coefficient of a value greater than 0.5 were selected as input in the model. However, an exception was applied to those related to O₃ where the value of Pearson correlation coefficient greater than or equal to 0.4 was used. Apart from the physical significance of the meteorological variable, such as sunshine rate to the production of O₃, this exception was necessary to prevent having too few inputs in the model that may fail to account for the fluctuation of O₃ levels. The available input variables included in the regression model is summarized in Table 1.

Table 2 shows the Pearson correlation coefficient between pollutants in different time series. The three-time series of SPM, SO₂, and O₃, at or even fall below, the level of significance of 0.5 after lag 2. Hence, in order to improve the accuracy of models and minimize the operation of models, only air pollutants and meteorological data at current day and previous day were used in this study to predict air pollutants level at the following day.

On the other hand, wind direction W is only available in the form of general directions, such as N, NE, and E. Therefore, Pearson correlation coefficient cannot be applied to find the relationship between air pollutants and wind direction. However, as mentioned above, wind direction is closely associated with the presence and dispersion of pollutants. Hence, wind direction is also selected as the available input variable in this study. The wind direction is separated into 16 discrete directions {N, NNE, NE, ENE, E, ...}. After applying corresponding analysis, it was found that only 7 out of the 16 wind directions are related to pollutant levels, namely, {N, NNE, E, ESE, SE, NW, NNW}. To represent these directions, a Boolean variable $W_i \in \{0, 1\}$ was used for each of them, where $i = 1$ to 7, rather than a number $W_i \in \{1, 2, \dots, 7\}$, so that no bias is incurred.

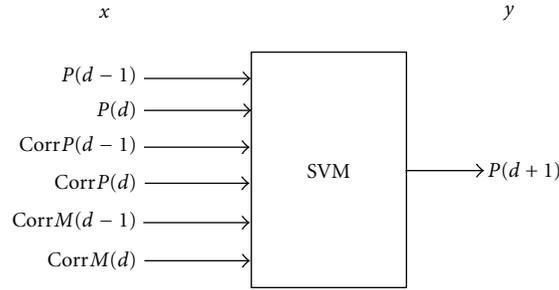


FIGURE 3: Data representation for pollutant predictive model.

TABLE 3: The training data and test data in each experiment.

Experiment	Training data	Test data
1-year experiment	2003–2005	2006 (a whole year)
Winter experiment	2003–2005	January 2006
Summer experiment	2003–June 2006	July 2006

In addition, since most of air pollutants are dissolvable, rainfall may be a critical impact feature to the output. However, after applying rainfall into modeling, it is found that the influence of rainfall is very low for the accuracy of models. Hence, rainfall was not selected as input variable in this study.

3. Methodology

3.1. Support Vector Machines. Support vector machines (SVMs) are known as an excellent tool for classification and regression problems [17–19], producing good generalization. The basic principle of SVM applies linear model to convert nonlinear class boundaries through some nonlinear mapping of the input vector into the high-dimensional feature space. Details of the working concept of SVM can be found in [13].

3.2. Kernel Selection. Kernel selection is a crucial issue for support vector machines. A kernel introduces nonlinearity into the SVM problem by mapping new input data, X , implicitly into Hilbert space via a function Φ where it may then be linearly separable. Since SVM only requires inner products of the nonlinearly mapped features $\Phi(X)$, a kernel becomes an efficient way to compute such an inner product and provides the same scalar output $k(X, X_t) = \Phi(X)^T \Phi(X_t)$, where k is a predefined kernel and X_t is the support vector. Different kernels will accommodate different nonlinear mapping and the performance of the resulting SVM is often hinged on the appropriate choice of the kernel [20]. Several kernels are commonly used in SVM for regression. These kernels including *Linear*, *Polynomial*, *Radial Basis Function (RBF)*, *Sigmoid*, and *Wavelet* were used

in this study to build SVM models as comparison. In general, these kernel functions are listed as follows, where $X, X_t \in R^m$:

$$\text{linear } k(X, X_t) = X^T \cdot X_t, \quad (1)$$

$$\text{polynomial } k(X, X_t) = (X^T \cdot X_t + 1)^n, \quad (2)$$

$$\text{RBF } k(X, X_t) = \exp\left(\frac{-\|X - X_t\|^2}{2\sigma^2}\right), \quad (3)$$

$$\text{sigmoid } k(X, X_t) = \tanh(X^T \cdot X_t + 1), \quad (4)$$

$$\text{wavelet } k(X, X_t) = \prod_{i=1}^m \varphi\left(\frac{X - X_t}{\sigma}\right). \quad (5)$$

In (5), φ can be any mother wavelet. In this study, *Morlet* function was selected.

4. Experiment Workflow

4.1. Data Sampling. The raw air pollutants and meteorological data from year 2003 to year 2006 were obtained from the website of DSMG (http://www.smg.gov.mo/www/e_index.php (last access: March 2012)). These data were divided into three groups for three experiments as shown in Table 3.

4.2. Data Normalization. Prior to modelling, it is necessary to normalize all selected features into same range to avoid the domination by any feature with large values. This normalization process leads to more stable and accurate predicted results. The features in training data and test data were normalized by subtracting and dividing by the feature means, that is,

$$\mathbf{x}_i \leftarrow \frac{\mathbf{x}_i - \bar{\mathbf{x}}_i}{\bar{\mathbf{x}}_i}, \quad (6)$$

where $\bar{\mathbf{x}}_i$ is the mean of the i th parameter of \mathbf{x} .

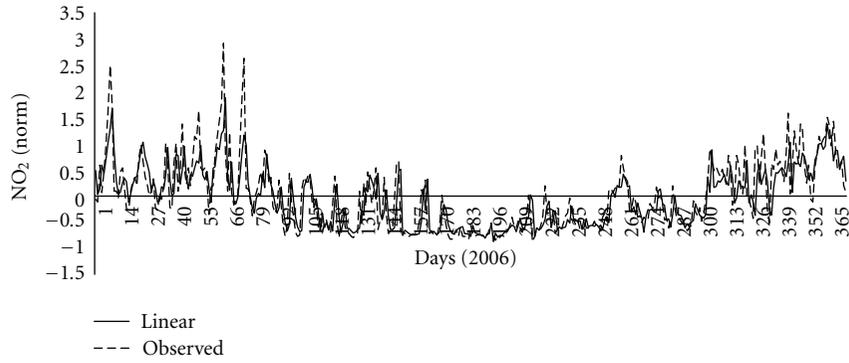
4.3. Modeling and Data Representation. As mentioned in Section 2, air pollutants and meteorological data at previous day and current day were used in this study to predict air pollutant level at the following day. In order to apply SVM for pollutant level forecasting, the representation of a pollutant

TABLE 4: The tested results of SPM, SO₂, NO₂, and O₃ in different kernel models in 1-year experiment.

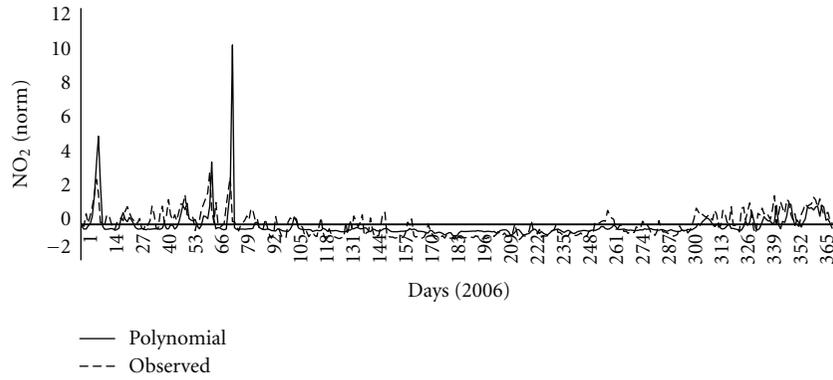
	Linear	Polynomial	Kernel RBF	Sigmoid	Wavelet
SPM					
MAE	0.312	0.421	0.310	1.459	0.457
RMSE	0.440	0.628	0.433	5.526	0.625
CWIA	16.7%	40.1%	16.0%	90.0%	26.9%
RE	18.9%	24.1%	17.8%	35.6%	31.8%
SO ₂					
MAE	0.513	0.663	0.515	13.072	0.679
RMSE	0.926	1.370	0.905	40.806	1.087
CWIA	18.4%	32.6%	17.9%	96.2%	23.8%
RE	7.4%	10.4%	8.5%	71.0%	8.5%
NO ₂					
MAE	0.267	0.408	0.265	1.365	0.429
RMSE	0.382	0.743	0.376	5.132	0.594
CWIA	11.2%	37.8%	10.5%	98.4%	24.1%
RE	12.9%	20.5%	12.3%	36.7%	27.9%
O ₃					
MAE	0.305	0.329	0.311	1.563	0.454
RMSE	0.403	0.434	0.408	1.921	0.579
CWIA	12.6%	15.8%	13.5%	58.3%	22.4%
RE	21.9%	25.8%	22.7%	84.4%	39.5%

TABLE 5: The tested results of SPM, SO₂, NO₂, and O₃ in different kernel models in winter experiment.

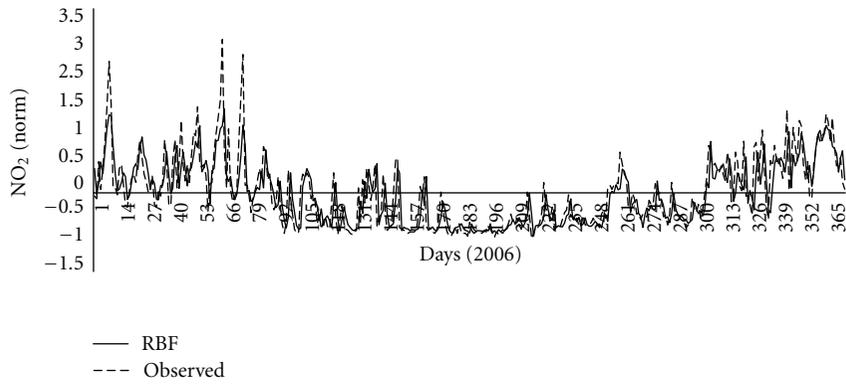
	Linear	Polynomial	Kernel RBF	Sigmoid	Wavelet
SPM					
MAE	0.307	0.335	0.317	3.724	0.493
RMSE	0.397	0.489	0.425	10.392	0.606
CWIA	17.6%	20.5%	19.8%	90.3%	36.8%
RE	25.8%	29.0%	32.3%	45.2%	61.3%
SO ₂					
MAE	0.823	1.024	0.811	32.112	1.126
RMSE	1.038	1.481	1.056	81.842	1.422
CWIA	17.5%	19.0%	17.9%	97.0%	33.0%
RE	32.3%	35.5%	35.5%	93.5%	51.6%
NO ₂					
MAE	0.293	0.491	0.305	3.736	0.459
RMSE	0.391	0.807	0.375	9.950	0.624
CWIA	17.6%	26.0%	15.3%	102.2%	39.6%
RE	25.8%	41.9%	16.1%	71.0%	41.9%
O ₃					
MAE	0.291	0.319	0.307	1.799	0.462
RMSE	0.369	0.388	0.369	2.217	0.571
CWIA	24.0%	27.7%	26.4%	76.1%	39.0%
RE	45.2%	51.6%	48.4%	93.5%	71.0%



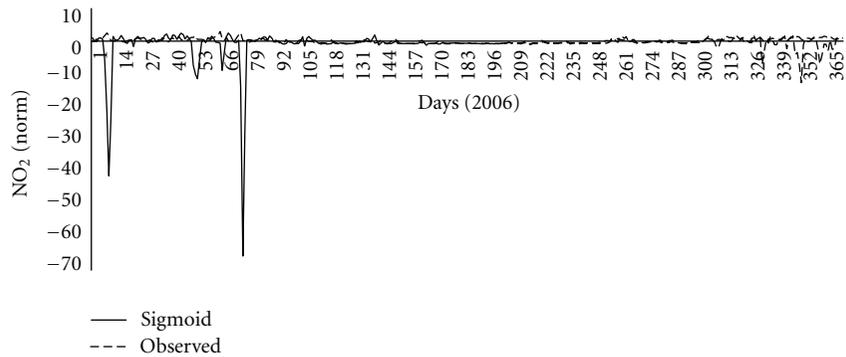
(a)



(b)



(c)



(d)

FIGURE 4: Continued.

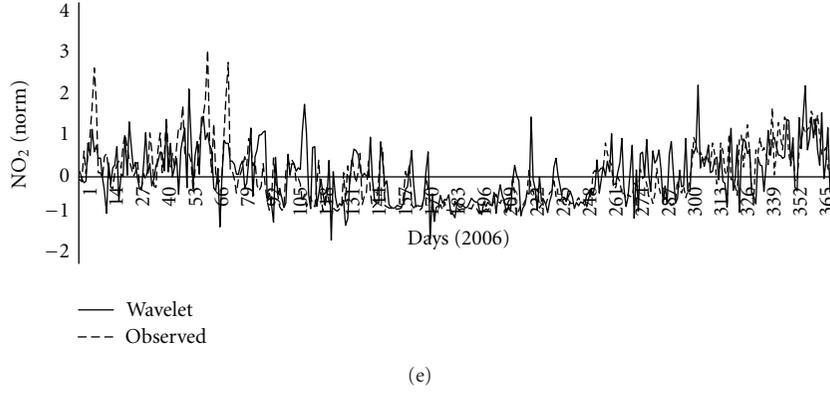


FIGURE 4: The predicted and observed levels of NO_2 using different kernel models in 1-year experiment.

level is defined as a pair (\mathbf{x}, y) . Generally the following features for a specific pollutant $P \in \{\text{SPM}, \text{SO}_2, \text{NO}_2, \text{O}_3\}$ are chosen for the representation of \mathbf{x} :

- (i) pollutant level at previous day: $P(d-1)$
- (ii) pollutant level at current day: $P(d)$
- (iii) correlated pollutant level at previous day: $\text{Corr}P(d-1)$
- (iv) correlated pollutant level at current day: $\text{Corr}P(d)$
- (v) correlated meteorological level at previous day: $\text{Corr}M(d-1)$
- (vi) correlated meteorological level at current day: $\text{Corr}M(d)$.

For example, if pollutant $P = \text{NO}_2$, then according to Table 1, its correlated pollutants $\text{Corr}P = \{\text{SPM}, \text{SO}_2\}$ and the correlated meteorological parameters $\text{Corr}M = \{W, T, \text{Hum}\}$, denoting the levels of SPM, SO_2 , wind direction, temperature, and humidity at previous day and current day, respectively. The representation of \mathbf{x} is then defined as

$$\begin{aligned} \mathbf{x} = & \langle \text{NO}_2(d-1), \text{NO}_2(d), \text{SPM}(d-1), \text{SPM}(d), \\ & \text{SO}_2(d-1), \text{SO}_2(d), W_i(d-1), W_i(d), \\ & T(d-1), T(d), \text{Hum}(d-1), \\ & \text{Hum}(d) \rangle \quad \text{for } i = 1 \text{ to } 7. \end{aligned} \quad (7)$$

Finally, the output $y = P(d+1)$ is the corresponding pollutant level of P (i.e., predicted pollutant level) at the following day.

This set of training data (\mathbf{x}, y) is then passed to SVM models. The concept is illustrated in Figure 3. For simplicity, the SVM models were named according to the kernel used in the model. Subsequently, five kinds of models for each pollutant in this study were as follows: *Linear model*, *Polynomial model*, *RBF model*, *Sigmoid model*, and *Wavelet model*. For $P \in \{\text{SPM}, \text{SO}_2, \text{NO}_2, \text{O}_3\}$, five modelling methods, three different experiments, and 60 different trained models were developed in total.

4.4. Experiment Environment. Modelling was performed on MATLAB 2007a platform where LIBSVM toolbox [21] and SVM Matlab toolbox [22] were employed to construct models. The hyperparameters (c and g) of SVM and the options of different kernels have been optimized.

5. Results

5.1. Error Measures. In order to effectively compare the accuracy among the models, four error measures were used in this study including mean absolute error (MAE), root mean squared error (RMSE), complementary Willmott's index of agreement (CWIA), and relative error (RE). It is necessary to set up the RE because in a warning system, attentions are usually focused on the level exceeding a particular dangerous level. The success of a forecasting system may be defined as whether the predicted value falls within an accepted error range relative to the true value [23]. In the following formulas, P_i and O_i represent the predicted level and observed level of i th day, respectively. O_{\max} and O_{\min} represent the maximum and minimum of observed level within each test set. n is number of data in the test sets:

$$\begin{aligned} \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |P_i - O_i|, \\ \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2}, \end{aligned} \quad (8)$$

$$\text{CWIA} = \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (|P'_i| - |O'_i|)^2},$$

where $P'_i = P_i - \bar{O}_i$, $O'_i = O_i - \bar{O}_i$, $\bar{O}_i = (1/n) \sum_{i=1}^n O_i$,

$$\text{RE} = \frac{1}{n} \sum_{i=1}^n E_i, \quad (9)$$

where

$$E_i = \begin{cases} 0 & \text{if } |P_i - O_i| < (O_{\max} - O_{\min}) \times 15\% \\ 1 & \text{otherwise.} \end{cases} \quad (10)$$

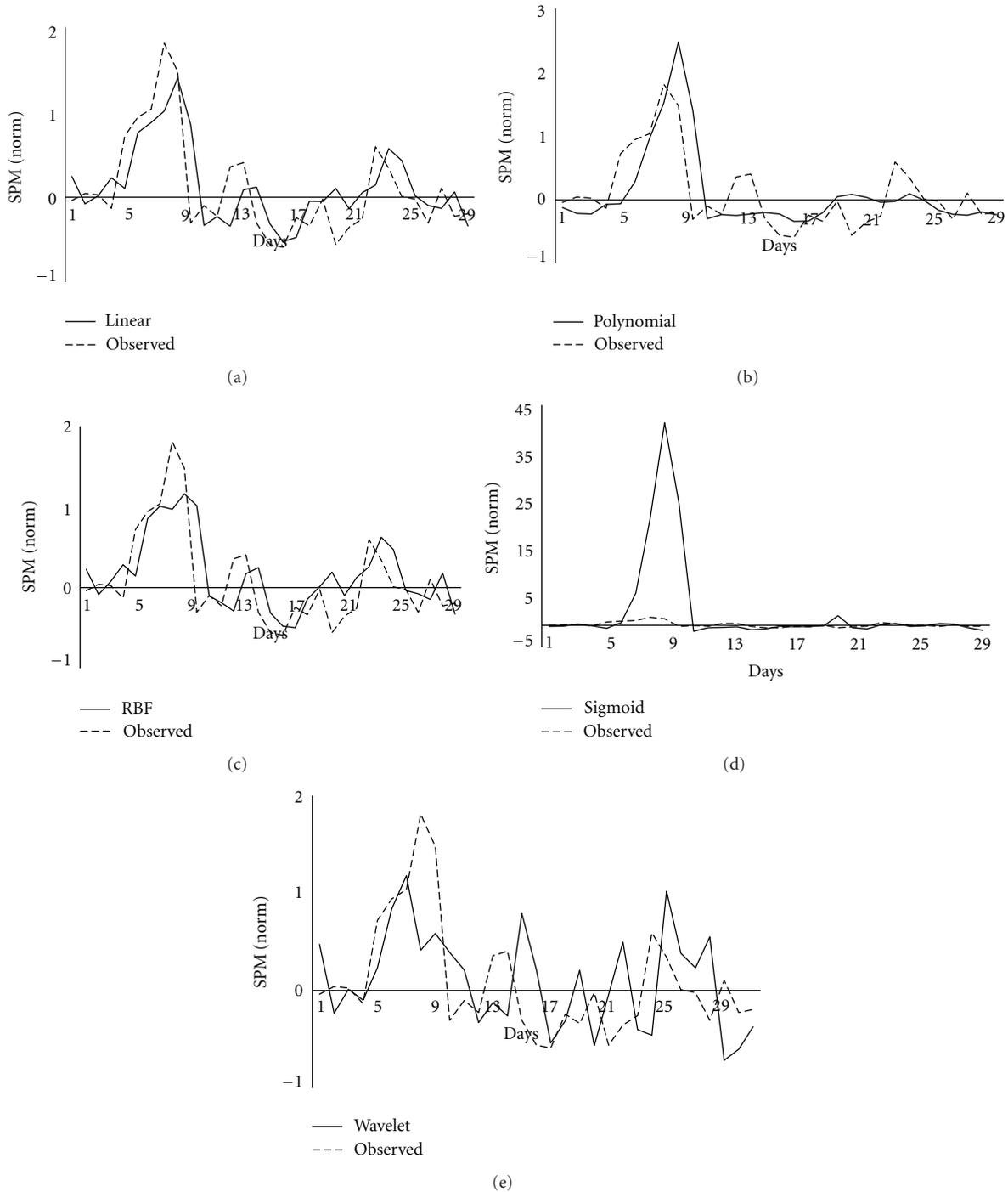


FIGURE 5: The predicted and observed levels of SPM using different kernel models in winter experiment (2006).

5.2. Prediction Results. Table 4 presents the results of SVM models under different kernels in 1-year experiment for SPM, SO₂, NO₂, and O₃. The bolded values indicate the best performance among the five testing models. *Linear model* and *RBF model* produced satisfactorily low errors for all pollutants. Moreover, the results of these two models were comparable. The results of *Polynomial model* and *Wavelet*

model were poor, evident in 3–18% higher than the results of *Linear model* and *RBF model*. *Sigmoid model* produced the highest errors. The predicted results of seasonal experiment showed the same pattern as in 1-year experiment (see Table 5 (winter experiment) and Table 6 (summer experiment)). From these results, *Linear model* and *RBF model* were capable of producing much higher generalization than other three

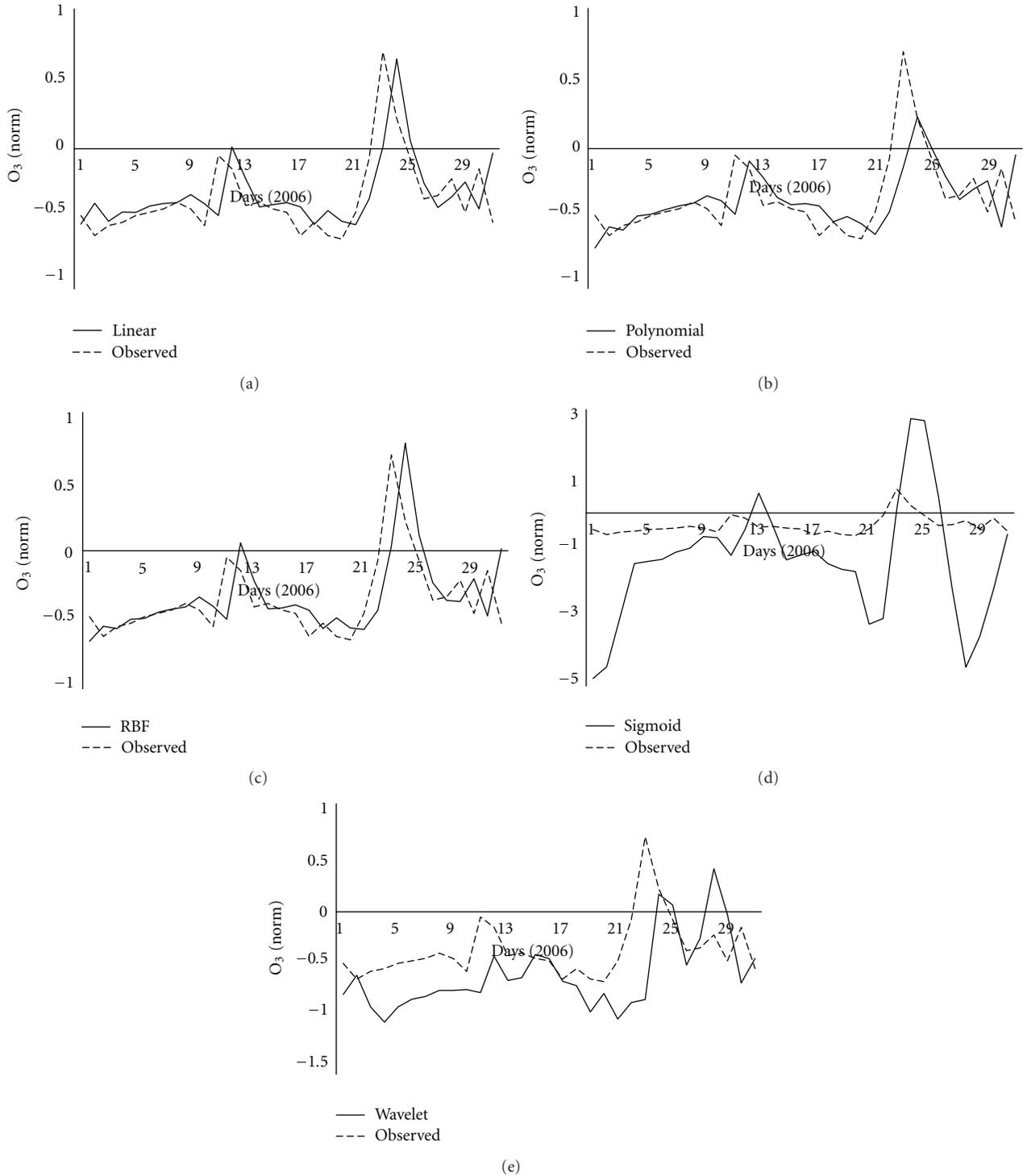


FIGURE 6: The predicted and observed levels of O_3 using different kernel models in summer experiment (2006).

models. In addition, the poor results of *Polynomial model* and *Sigmoid model* may be caused by the use of more hyperparameters, which are difficult to optimize.

5.3. Matching of Predicted and Observed Pollution Levels. The exemplary plots of predicted and observed level of

NO_2 in 1-year experiment, SPM in winter experiment, and O_3 in summer experiment are depicted in Figures 4 to 6, respectively. Although there were some lagging and underestimations, the predicted levels (Figure 4) produced by *Linear model* and *RBF model* followed the trend of observed level of NO_2 pretty well in 1-year experiment. However,

TABLE 6: The tested results of SPM, SO₂, NO₂, and O₃ in different kernel models in summer experiment.

	Kernel				
	Linear	Polynomial	RBF	Sigmoid	Wavelet
SPM					
MAE	0.165	0.233	0.169	0.438	0.258
RMSE	0.243	0.280	0.248	0.685	0.379
CWIA	34.8%	48.1%	34.1%	62.2%	41.5%
RE	29.0%	54.8%	32.3%	58.1%	45.2%
SO ₂					
MAE	0.069	0.221	0.058	0.748	0.111
RMSE	0.089	0.244	0.086	1.022	0.164
CWIA	9.3%	43.9%	8.8%	68.6%	24.1%
RE	48.4%	93.5%	32.3%	100.0%	61.3%
NO ₂					
MAE	0.653	0.719	0.675	3.646	0.578
RMSE	0.809	1.352	0.824	9.418	0.723
CWIA	76.7%	92.4%	77.1%	96.6%	78.6%
RE	67.7%	41.9%	67.7%	80.6%	64.5%
O ₃					
MAE	0.165	0.154	0.170	1.495	0.322
RMSE	0.227	0.237	0.243	1.932	0.435
CWIA	25.5%	30.3%	24.6%	77.5%	47.0%
RE	29.0%	25.8%	22.6%	90.3%	61.3%

the other three models failed to follow observed level at all. In winter experiment (Figure 5), *Linear model* performed the best, while the other four models failed to follow the trend of the observed level, especially *Sigmoid model*. In summer experiment (Figure 6), *Linear model*, *RBF model*, and *Polynomial model* showed good performances. However, *Polynomial model* cannot match the peaks of the observed levels. Both *Sigmoid model* and *Wavelet model* produced poor prediction comparing to other three models. It is clear that *Linear model* and *RBF model* performed the best and their predicted results were the closest to the observed levels, regardless in the 1-year experiment or the seasonal experiment.

6. Conclusion

Using observed meteorological and pollutant data, SVM models for forecasting daily ambient air pollutant were constructed. The prediction results of *Linear model* and *RBF model* showed a relative good fit to the observed test set of over one year of data, particularly for SO₂ and NO₂. In seasonal experiment, *Linear model* and *RBF model* also outperformed other three tested models, although some lagging and underestimations of these two models occurred in winter experiment. Comparing to these five studied models, it was evident that using *Linear kernel* and *RBF kernel* in SVM model for air pollutant forecasting in Macau produced superior results with relatively lower errors. It is believed that SVM model with *Linear kernel* or *RBF*

kernel can also produce good performance for air pollutant forecasting in other similar developing cities, or even other time series prediction in similar situation.

Although *Linear model* and *RBF model* outperformed other three tested models, both of these two models still suffer underestimation of high levels of pollutants. How to solve this problem to improve the accuracy of prediction model is the future works. Some literature [24] attempted to integrate discrete wavelet transform (DWT) with SVM for a higher accuracy. Hence, we will attempt to integrate other machine learning methods, for example, genetic algorithm (GA), with SVM to improve the accuracy and efficiency of model in the future.

Acknowledgment

The research was sponsored by research committee, University of Macau, under Grant no. MYRG141(Y2-L2)-FST11-IWF.

References

- [1] J. A. Bernstein, N. Alexis, C. Barnes et al., "Health effects of air pollution," *Journal of Allergy and Clinical Immunology*, vol. 114, no. 5, pp. 1116–1123, 2004.
- [2] A. Seaton, W. MacNee, K. Donaldson, and D. Godden, "Particulate air pollution and acute health effects," *Lancet*, vol. 345, no. 8943, pp. 176–178, 1995.
- [3] C. A. Pope III, D. V. Bates, and M. E. Raizenne, "Health effects of particulate air pollution: time for reassessment?"

- Environmental Health Perspectives*, vol. 103, no. 5, pp. 472–480, 1995.
- [4] R. McConnell, K. Berhane, F. Gilliland et al., “Air pollution and bronchitic symptoms in Southern California children with asthma,” *Environmental Health Perspectives*, vol. 107, no. 9, pp. 757–760, 1999.
- [5] WHO, “WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide,” 2005.
- [6] DSEC, “Population Estimate of Macao,” 2008, http://www.dsec.gov.mo/getAttachment/9cc40c96-fa67-400b-b66f-d77a-f509754a/E_POP.FR.2008_Y.aspx.
- [7] DSEC, “Macao Resident Population Projections 2007–2031,” 2008, http://www.dsec.gov.mo/getAttachment/e92b4427-6057-491a-80b7-f4255b65f1ca/E_PPRM.PUB.2007_Y.aspx.
- [8] DSEC, “Transport and Communications Statistics 2009,” 2009, http://www.dsec.gov.mo/getAttachment/7c480491-15cf-4018-a0b8-19fa951be8a2/E_ETC.FR.2009_M12.aspx.
- [9] S. L. Reich, D. R. Gomez, and L. E. Dawidowski, “Artificial neural network for the identification of unknown air pollution sources,” *Atmospheric Environment*, vol. 33, no. 18, pp. 3045–3052, 1999.
- [10] M. W. Gardner and S. R. Dorling, “Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences,” *Atmospheric Environment*, vol. 32, no. 14–15, pp. 2627–2636, 1998.
- [11] M. Kolehmainen, H. Martikainen, and J. Ruuskanen, “Neural networks and periodic components used in air quality forecasting,” *Atmospheric Environment*, vol. 35, no. 5, pp. 815–825, 2001.
- [12] W. Z. Lu, H. Y. Fan, A. Y. T. Leung, and J. C. K. Wong, “Analysis of pollutant levels in central Hong Kong applying neural network method with particle swarm optimization,” *Environmental Monitoring and Assessment*, vol. 79, no. 3, pp. 217–230, 2002.
- [13] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1st edition, 1995.
- [14] W. Z. Lu and W. J. Wang, “Potential assessment of the “support vector machine” method in forecasting ambient air pollutant trends,” *Chemosphere*, vol. 59, no. 5, pp. 693–701, 2005.
- [15] I. Juhos, L. Makra, and B. Tóth, “Forecasting of traffic origin NO and NO₂ concentrations by Support Vector Machines and neural networks using Principal Component Analysis,” *Simulation Modelling Practice and Theory*, vol. 16, no. 9, pp. 1488–1502, 2008.
- [16] W. C. Wang, K. W. Chau, C. T. Cheng, and L. Qiu, “A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series,” *Journal of Hydrology*, vol. 374, no. 3–4, pp. 294–306, 2009.
- [17] P. K. Wong, C. M. Vong, L. M. Tam, and K. Li, “Data preprocessing and modelling of electronically-controlled automotive engine power performance using kernel principal components analysis and least squares support vector machines,” *International Journal of Vehicle Systems Modelling and Testing*, vol. 3, no. 4, pp. 312–330, 2008.
- [18] P. K. Wong, L. M. Tam, K. Li, and C. M. Vong, “Engine idle-speed system modelling and control optimization using artificial intelligence,” *Proceedings of the Institution of Mechanical Engineers D*, vol. 224, no. 1, pp. 55–72, 2010.
- [19] C. M. Vong and P. K. Wong, “Engine ignition signal diagnosis with Wavelet Packet Transform and Multi-class Least Squares Support Vector Machines,” *Expert Systems with Applications*, vol. 38, no. 7, pp. 8563–8570, 2011.
- [20] T. Jebara, “Multi-task feature and kernel selection for SVMs,” in *21st International Conference on Machine Learning (ICML '04)*, pp. 433–440, Banff, Canada, July 2004.
- [21] C. C. Chang and C. J. Lin, “LIBSVM—A Library for Support Vector Machines,” <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [22] “SVM and Kernel Methods Matlab Toolbox,” <http://asi.insa-rouen.fr/enseignants/~arakotom/toolbox/index.html>.
- [23] W. F. Ip, C. M. Vong, J. Y. Yang, and P. K. Wong, “Least squares support vector prediction for daily atmospheric pollutant level,” in *9th IEEE/ACIS International Conference on Computer and Information Science (ICIS '10)*, pp. 23–28, August 2010.
- [24] S. Osowski and K. Garanty, “Forecasting of the daily meteorological pollution using wavelets and support vector machine,” *Engineering Applications of Artificial Intelligence*, vol. 20, no. 6, pp. 745–755, 2007.

Research Article

A Memory-Based Hysteresis Model in Piezoelectric Actuators

Guilin Zhang,¹ Chengjin Zhang,¹ and Jason Gu^{1,2}

¹ School of Control Science and Engineering, Shandong University, Jinan 250061, Shandong, China

² Electrical and Computer Engineering, Dalhousie University, Halifax, NS, Canada

Correspondence should be addressed to Chengjin Zhang, cjzhang@sdu.edu.cn

Received 12 January 2012; Revised 25 March 2012; Accepted 5 April 2012

Academic Editor: Pak Kin Wong

Copyright © 2012 Guilin Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A mathematical memory-based model is proposed to capture the hysteresis behavior in piezoelectric actuators. It is observed that the ascending (descending) hysteresis curves are alike and converge to one point without memory saturation. Therefore, two dominant curves are determined and expressed as continuous functions, and the other hysteresis curves are modeled using two dominant curves through nonlinear transforming of coordinate axis. In the event of memory saturation, a new converging point is used to compensate the model prediction error. The experimental study has been carried out and our proposed model prediction method is compared with PI model and the linear model. It shows that the proposed model prediction method is better than other two methods.

1. Introduction

Piezoelectric ceramics are widely used as actuators in nano-/micropositioning mechatronic systems due to their fast frequency response, nanometer scale resolution, and high stiffness. Since the materials of piezoelectric actuators are ferroelectric, nonlinear hysteresis behavior is commonly observed in such actuators in response to an applied electric field. Thus, it is very challenging to design high-performance servo controller for nano-/micropositioning mechatronic systems [1].

In the past few years, many studies on the compensation for the positioning of the piezoelectric actuators have been reported. One simple useful method is the model-based feed-forward control scheme. This method cascades an inverse hysteresis model in series with an actuator plant to cancel out the effect of nonlinearity and achieve a relatively linear response. The more precision the inverse hysteresis model is, the better the control scheme tracking performance is. Consequently, modeling hysteresis is the first step to track control of piezoelectric actuators.

Numerous research works have been done to model the hysteresis nonlinearity of piezoelectric actuators. The most well-known hysteresis model is the Preisach model [2–11]. It is the broadly used approach in modeling and compensation of hysteresis in piezoelectric actuators. However, Preisach

model does not have an analytical inverse. Ge and Jouaneh improved classical Preisach model that can predict the hysteresis response of a piezoelectric actuator driven by a periodic sinusoidal or triangular input signal [5] and developed a computer-based tracking control approach for piezoelectric actuators based on linearizing the hysteresis nonlinearity [6]. Tan et al. formulated and proposed the value inversion algorithm using a class of discretized Preisach operators [9]. They proposed an adaptive identification of hysteresis in smart materials and developed an adaptive inverse control scheme to update the weight of Preisach operators [10].

Another popular hysteresis model is Prandtl-Ishlinskii (PI) model, which is a subclass of the Preisach model. The advantage of PI model is that its inverse is also a PI type with different threshold and weighting values. Kuhnen developed an inverse scheme based on the inverse scheme of PI model [12, 13]. However, the PI operator in PI model has a symmetry property around the center of the loop, while hysteresis response of a piezoelectric actuator is not symmetric in practice. Bashash and Jalili investigated the nature of hysteresis in piezoelectric materials and discussed a modified PI model that describes the asymmetric and residual displacement properties simultaneously [14]. Jiang et al. proposed another modified PI model based on two asymmetric operators to model the asymmetric property of hysteresis [15]. The other disadvantage of PI model is

that it cannot compensate for the saturated hysteresis. Al Janaideh Mohammad et al. proposed a generic PI model based on a generalized play operator with different loading and unloading envelop functions in order to characterize asymmetric and saturated hysteresis nonlinearities [16]. They also proposed the modified generalized PI model that has the exact analytical inverse [17].

Xu and Wong applied least squares support vector machine (LS-SVM) to capture the rate-independent and rate-dependent hysteresis nonlinearities [18, 19]. Moreover, a set of models are proposed to describe the hysteresis properties including targeting turning points, curve alignment and wiping-out property. Tzen et al. applied an exponential curve to fit the hysteresis path when the piezoelectric actuator operated away from the saturation [20]. Sun et al. proposed a hysteresis model based on similarities of the hysteresis curves and the turning points [21]. Bashash and Jalili disclosed the memory-dominant nature of hysteresis in piezoelectric materials and developed a memory-based mathematical hysteresis model [22–24]. They adopted the proposed model in an inverse model-based control scheme for feedforward compensation of hysteresis nonlinearity [22].

In this paper, we adopt the mathematical transformation to describe the hysteresis curve functions. The ascending and descending loading curves can be determined and expressed as two continuous functions. Then the rest of hysteresis curves adopt their shape from the dominant curves. In addition, a new converging is updated to compensate the model prediction error in the event of memory saturation. It is demonstrated that the proposed model prediction is noticeably improved compared with the model without memory operator.

The remainder of this paper is organized as follows. First, a mathematical memory-based hysteresis model in piezoelectric actuators is proposed in Section 2. Moreover, the approximate congruency property of the proposed model is validated in this section. Then the experimental setup and the proposed model performances are given in Section 3. In Section 4, a new converging point is updated to address the memory saturation. Finally, some concluding remarks are provided in Section 5.

2. Hysteresis Model

Hysteresis trajectory starts moving on the ascending loading curve as depicted in Figure 1. This curve can be approximated by a monotonically increasing continuous function. As shown in Figure 1, the ascending curves are alike and converge to the upper targeting point. When the direction of the input changes, the trajectory breaks its path and moves downward on the descending loading curve that can be approximated by another monotonically increasing continuous function. All the descending curves are also alike and converge to the lower converging point. The ascending and descending dominant curves can be expressed as two monotonically continuous functions, $f_{ra}(v)$ and $f_{rd}(v)$, respectively. In addition, the rest of hysteresis curves can adopt their shape from these dominant curves.

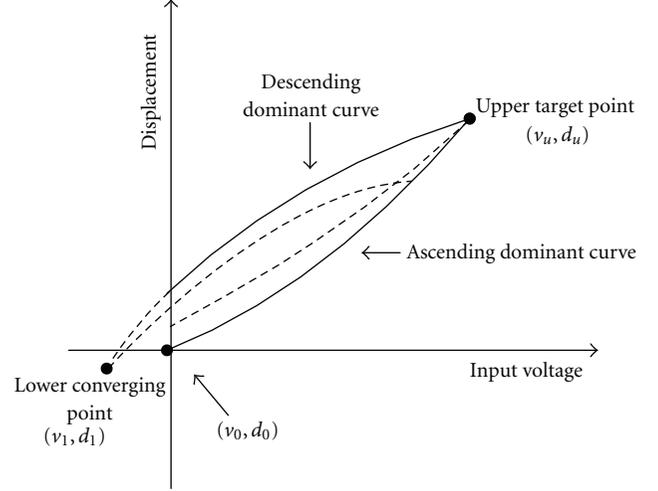


FIGURE 1: Ascending and descending dominant curves.

For any ascending trajectory starting from point (v_1, d_1) and (v_2, d_2) , the following linear function is obtained:

$$d_a(v) = k_a f_{ra}(v) + b_a, \quad (1)$$

where $d_a(v)$ represents the ascending hysteresis trajectory between points (v_1, d_1) and (v_2, d_2) , when the input voltage v varies between v_1 and v_2 , k_a and b_a are given by

$$k_a = \frac{d_u - d_1}{f_{ra}(v_u) - f_{ra}(v_0)}, \quad b_a = d_1 - k_a * f_{ra}(v_0), \quad (2)$$

where v_0 and v_u are the initial voltage and the upper converging voltage of the dominant curve depicted in Figure 1.

Replacing k_a and b_a from (2) into (1) yields

$$\begin{aligned} d_a(v, v_1, d_1, v_2, d_2) \\ = d_1 + \frac{d_u - d_1}{f_{ra}(v_u) - f_{ra}(v_0)} \\ \times (f_{ra}(v) - f_{ra}(v_0)). \end{aligned} \quad (3)$$

Once the initial voltage v_1 and displacement d_1 are decided, the hysteresis curves are obtained by using the dominant curve. However, when the initial voltage v_1 substitute v in (3),

$$\begin{aligned} d_a(v_1) \\ = d_1 + \frac{d_u - d_1}{f_{ra}(v_u) - f_{ra}(v_0)} (f_{ra}(v_1) - f_{ra}(v_0)) \neq d_1. \end{aligned} \quad (4)$$

Then, the ascending curve function is modified as

$$\begin{aligned} d_a(v) \\ = d_1 + \frac{d_u - d_1}{f_{ra}(v_u) - f_{ra}(v_0)} (f_{ra}(mv + (1 - m)v_u) - f_{ra}(v_0)), \end{aligned} \quad (5)$$

where $m = (v_u - v_0)/(v_u - v_1)$ and when $v = v_1$, $d_a(v_1) = d_1$ and when $v = v_u$, $d_a(v_u) = d_u$. It demonstrates that the model can start from the initial displacement and reach the upper converging point.

Similarly, when the input voltage v varies between v_1 and v_2 , the descending curve function is given by

$$d_d(v) = k_d f_{rd}(nv + (1 - n)) + b_d, \quad (6)$$

where $d_d(v)$ represents the descending hysteresis trajectory between points (v_1, d_1) and (v_2, d_2) , k_d and b_d are the same as k_a and b_a and given by

$$k_d = \frac{d_1 - d_1}{f_{rd}(v_1) - f_{rd}(v_u)}, \quad (7)$$

$$b_d = d_1 - k_d * f_{rd}(v_u), \quad n = \frac{v_1 - v_u}{v_1 - v_1},$$

where (v_u, d_u) is the upper converging point and v_u is also the initial voltage of the descending dominant curves, (v_l, d_l) is the lower converging point and can be identified. Then (6) can be transformed into

$$d_d(v, v_1, d_1, v_2, d_2)$$

$$= d_1 + \frac{d_l - d_1}{f_{rd}(v_l) - f_{rd}(v_u)} (f_{rd}(nv + (1 - n)v_l) - f_{rd}(v_u)). \quad (8)$$

The congruency property states that any two minor hysteresis loops are identical and have the same shape if they are generated by back-and-forth variations of the input between two identical extrema. Here, we consider the case of two points (v_x, d_{x1}) and (v_x, d_{x2}) ($d_{x1} \neq d_{x2}$). When the input voltage reaches v_y , the displacement d_{y1} and d_{y2} are obtained by using (5).

$$d_{y1}$$

$$= d_{x1} + \frac{d_u - d_{x1}}{f_{ra}(v_u) - f_{ra}(v_0)} (f_{ra}(mv + (1 - m)v_l) - f_{ra}(v_0)), \quad (9)$$

$$d_{y2}$$

$$= d_{x2} + \frac{d_u - d_{x2}}{f_{ra}(v_u) - f_{ra}(v_0)} (f_{ra}(mv + (1 - m)v_l) - f_{ra}(v_0)). \quad (10)$$

Subtracting (9) from (10) yields

$$d_{y2} - d_{y1}$$

$$= d_{x2} + \frac{d_u - d_{x2}}{f_{ra}(v_u) - f_{ra}(v_0)}$$

$$\times (f_{ra}(mv_y + (1 - m)v_u) - f_{ra}(v_0))$$

$$- d_{x1} + \frac{d_u - d_{x1}}{f_{ra}(v_u) - f_{ra}(v_0)}$$

$$\times (f_{ra}(mv_y + (1 - m)v_u) - f_{ra}(v_0))$$

$$= (d_{x2} - d_{x1}) \left(1 + \frac{f_{ra}(mv_y + (1 - m)v_u) - f_{ra}(v_0)}{f_{ra}(v_u) - f_{ra}(v_0)} \right)$$

$$= (d_{x2} - d_{x1})(1 + \sigma). \quad (11)$$

When $v_y \ll v_u$, $\sigma \approx 0$, then $d_{y2} - d_{y1} \approx (d_{x2} - d_{x1})$. That is to say, the congruency property of the proposed model can be approximately satisfied only under the condition that the input voltage is far away from the upper converging point.

The implementation of the model needs to identify the ascending and descending dominant curves. Some structures have been proposed for these curves, including second-order polynomials, third-order polynomials, and exponential functions. Here, two-third-order polynomials are used for the approximation of the ascending and descending dominant curves. The polynomials are expressed as

$$f_{ra}(v) = \sum_{i=0}^3 a_i v^{3-i} = a_0 v^3 + a_1 v^2 + a_2 v + a_3, \quad (12)$$

$$f_{rd}(v) = \sum_{i=0}^3 d_i v^{3-i} = d_0 v^3 + d_1 v^2 + d_2 v + d_3,$$

where a_i and d_i are the dominant curve coefficients that can be identified by utilizing the least square identification method.

The proposed model has several advantages compared to the Preisach model. Firstly, it only needs eight parameters compared with the Preisach model which needs a large number of parameters. Secondly, the numerical inversion of the proposed model can be easily obtained in real time based on the mathematical formulation.

3. Experimental Setup and Model Verification

To investigate the effectiveness of the proposed piezoelectric actuator hysteresis model, a set of experiment on a PST150/7/40VS12 PZT-driven is demonstrated using the PC-based control system. Figure 2(a) shows the experimental block diagram. The setup includes a personal computer, a power controller, and a piezoelectric actuator, which provides maximum 40 μm displacement and an integrated high-resolution strain gauge position (SGS) sensor. The host computer generates the control codes, which are written in Visual C++ and run through the EPP (Enhanced Parallel Port). The digital signal is converted by 16-bit D/A converter and amplified by a high-voltage amplifier (PVC-150S1). The actual actuator output displacements are measured by a SGS sensor and converted to a digital signal by a 16-bit A/D converter. Then the displacement data are sent to the host computer. The experimental setup (without PC) is shown in Figure 2(b).

The dominant curves should be obtained first. A first-order reversal signal is applied to excite the piezoelectric actuator, which is shown in Figure 3(a). The proposed model is applied to predict the output displacement. Figure 3(b)

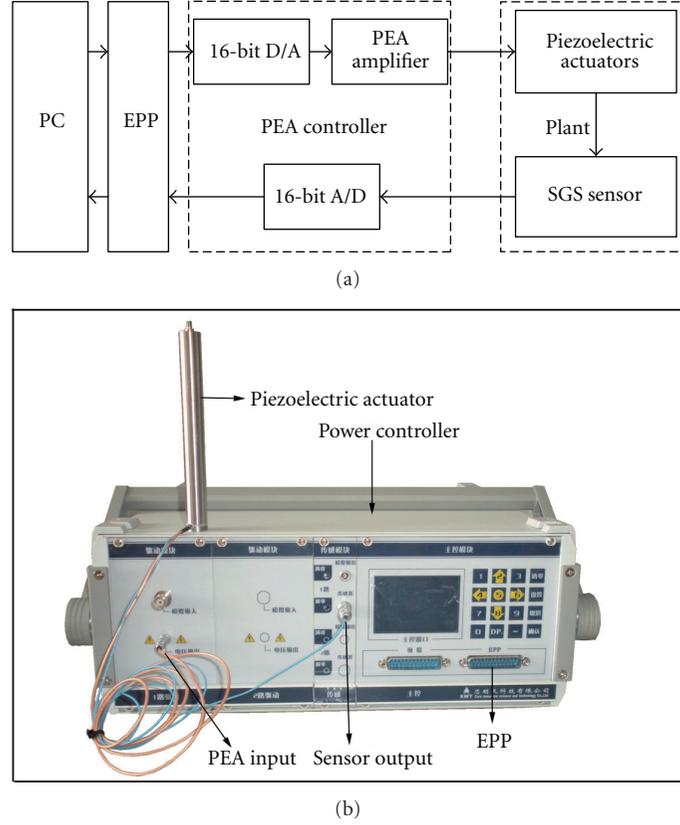


FIGURE 2: (a) The experimental block diagram. (b) the experimental setup.

depicts the experimental displacement and the proposed model prediction. The maximum and root-mean-square (rms) error percentages of the proposed model are 0.96% and 0.49%, respectively.

To demonstrate the effectiveness of the proposed model, the linear model and PI models are considered for comparison. Both the linear and PI model are simulated based on the same input given in Figure 3(a). The responses of these models are shown in Figures 3(c) and 3(d), respectively. The maximum and rms error percentages are 4.96% and 4.10% for linear model, and 1.32% and 0.52% for PI model, respectively. The modeling performance of the proposed model has been improved compared with the linear model and PI model.

4. Saturated Memory Model

The upper converging point is not a fixed point. In this section, the memory saturation is considered. The point 2 is a converging point as shown in Figure 4. Path 1-2 is the dominant ascending curve. When the input signal overtakes point 2, the new hysteresis trajectory will follow the path 2-5 that is the extended path 1-2. Meanwhile, point 5 becomes the new upper converging point after the input voltage reaches point 5. Consequently, when the input signal exceeds a converging point, the new hysteresis trajectory will converge to a new point.

We consider the case that the i_{th} ascending trajectory starts from point (v_{i1}, d_{i1}) and (v_{u1}, d_{u1}) , $v_{u1} > v_u$. Because the converging point has varied in the next ascending curves, then the next ascending curves in (5) can be updated to (13) as follows:

$$d_a(v) = d_1 + \frac{d_{u1} - d_1}{f_{ra}(v_{u1}) - f_{ra}(v_0)} \times (f_{ra}(mv + (1 - m)v_{u1}) - f_{ra}(v_0)). \quad (13)$$

Different from the ascending curves, the descending curves converge to a fixed point (v_l, d_l) . Thus, the descending curve functions keep unchanged in the event of memory saturation.

To demonstrate the effectiveness of the saturated memory model, a simulation study is carried out. The input profile shown in Figure 5(a) is applied to excite the piezoelectric actuator. The experimental and the model responses without memory operator are demonstrated in Figure 5(b). As seen from Figure 5(b), the proposed model response diverges from the correct trajectory when it hits the dominant maximum input voltage. Then the saturated memory model response is given in Figure 5(c), which demonstrates the performance enhancement of the saturated memory model.

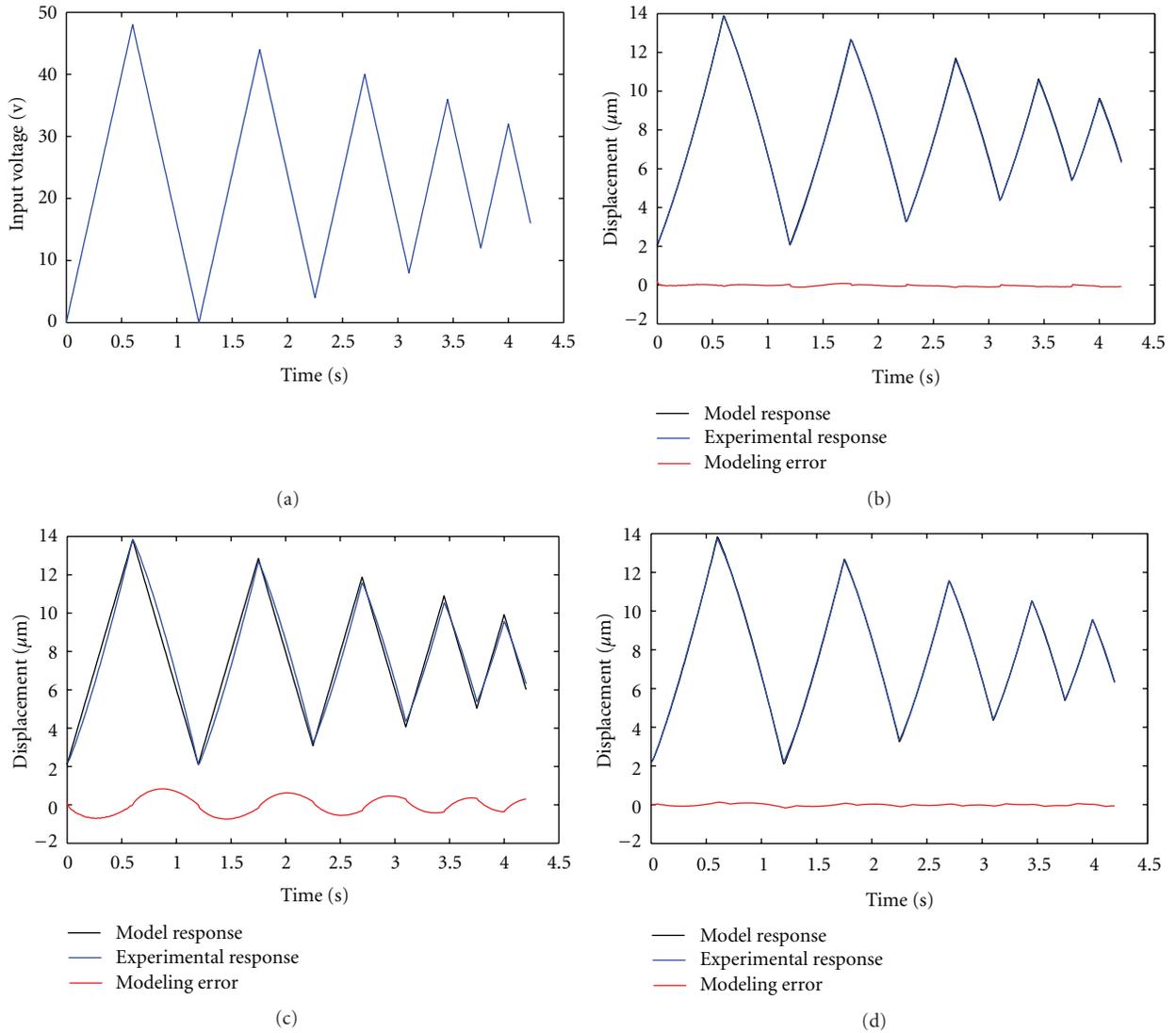


FIGURE 3: (a) The first-order reversal signal. (b) The experimental response and the proposed model response. (c) The experimental response and the linear model response, and (d) the experimental response and the PI model response.

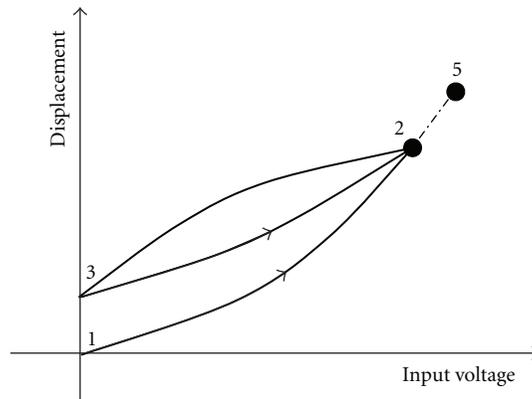


FIGURE 4: Hysteresis curve movement path.

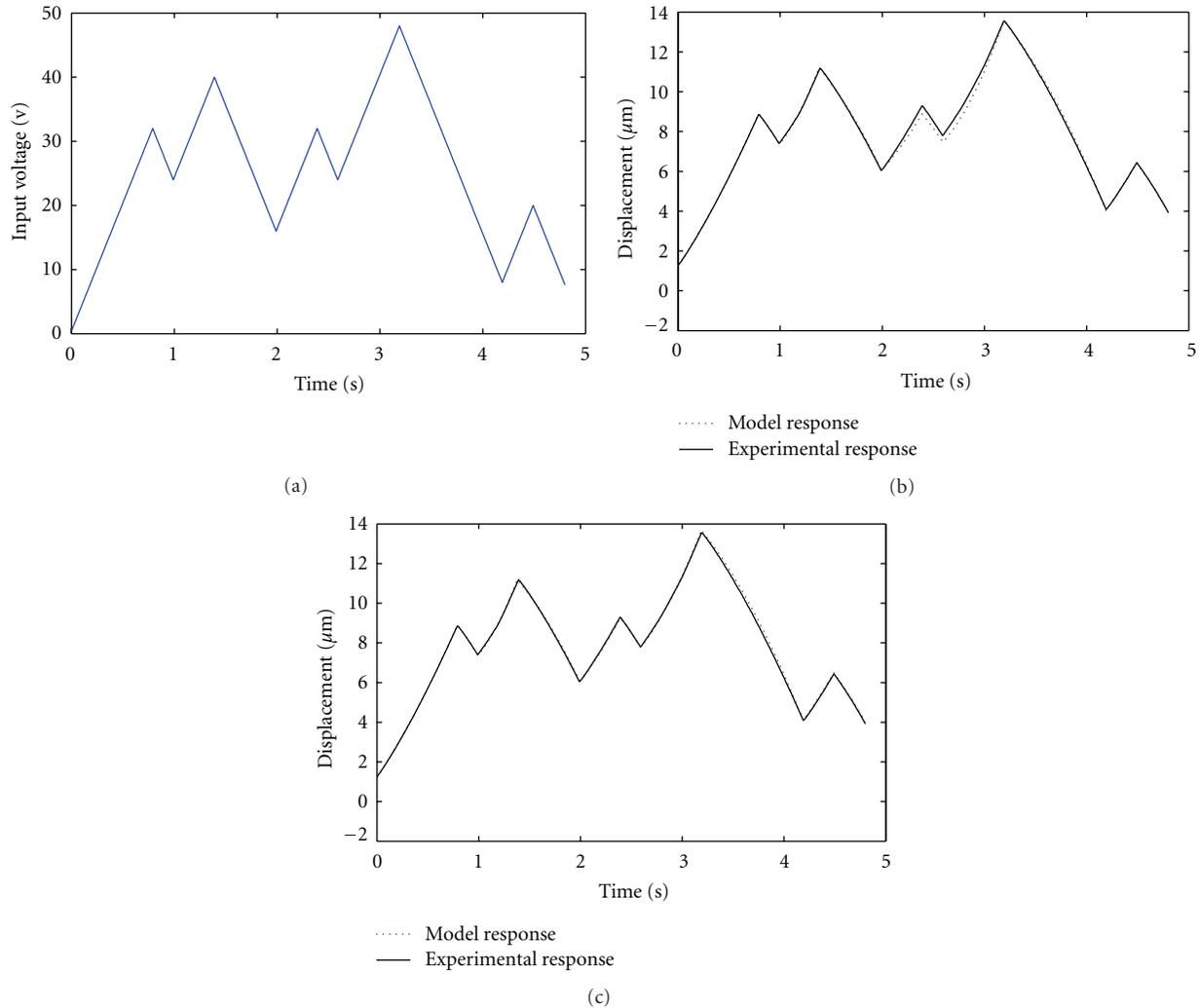


FIGURE 5: (a) The experimental input signal. (b) The experimental response and the proposed model response, and (c) the experimental response and the saturated memory model response.

5. Conclusions

Hysteresis is the main nonlinearity in piezoelectric actuators. Based on the similarities of the hysteresis trajectory, a new mathematical hysteresis model was proposed to capture the hysteresis behavior in this paper. Utilizing the mathematical transformation, a memory-based modeling framework was developed and experimentally validated on a piezoelectric actuator. Moreover, a new converging point was updated to compensate the prediction error when the hysteresis path hits the upper converging point. The experimental and simulation results demonstrate the effectiveness of the saturated memory model.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (61174044) and the Shandong Province Natural Science Foundation (ZR2010FM016).

References

- [1] S. Devasia, E. Eleftheriou, and S. O. R. Moheimani, "A survey of control issues in nanopositioning," *IEEE Transactions on Control Systems Technology*, vol. 15, no. 5, pp. 802–823, 2007.
- [2] A. A. Adly, I. D. Mayergoyz, and A. Bergqvist, "Preisach modeling of magnetostrictive hysteresis," *Journal of Applied Physics*, vol. 69, no. 8, pp. 5777–5779, 1991.
- [3] D. Croft, G. Shed, and S. Devasia, "Creep, hysteresis, and vibration compensation for piezoactuators: atomic force microscopy application," *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, vol. 123, no. 1, pp. 35–43, 2001.
- [4] C. Natale, F. Velardi, and C. Visone, "Identification and compensation of Preisach hysteresis models for magnetostrictive actuators," *Physica B*, vol. 306, no. 1–4, pp. 161–165, 2001.
- [5] P. Ge and M. Jouaneh, "Modeling hysteresis in piezoceramic actuators," *Precision Engineering*, vol. 17, no. 3, pp. 211–221, 1995.
- [6] P. Ge and M. Jouaneh, "Tracking control of a piezoceramic actuator," *IEEE Transactions on Control Systems Technology*, vol. 4, no. 3, pp. 209–216, 1996.

- [7] P. Ge and M. Jouaneh, "Generalized preisach model for hysteresis nonlinearity of piezoceramic actuators," *Precision Engineering*, vol. 20, no. 2, pp. 99–111, 1997.
- [8] X. Tan and J. S. Baras, "Modeling and control of hysteresis in magnetostrictive actuators," *Automatica*, vol. 40, no. 9, pp. 1469–1480, 2004.
- [9] X. Tan, J. S. Baras, and P. S. Krishnaprasad, "Control of hysteresis in smart actuators with application to micro-positioning," *Systems and Control Letters*, vol. 54, no. 5, pp. 483–492, 2005.
- [10] X. Tan and J. S. Baras, "Adaptive identification and control of hysteresis in smart materials," *IEEE Transactions on Automatic Control*, vol. 50, no. 6, pp. 827–839, 2005.
- [11] G. Song, J. Zhao, X. Zhou, and J. A. De Abreu-García, "Tracking control of a piezoceramic actuator with hysteresis compensation using inverse Preisach model," *IEEE/ASME Transactions on Mechatronics*, vol. 10, no. 2, pp. 198–209, 2005.
- [12] K. Kuhnen and H. Janocha, "Adaptive inverse control of piezoelectric actuators with hysteresis operators," in *Proceedings of the European Control Conference*, Karlsruhe, Germany, 1999, paper F 0291.
- [13] P. Krejci and K. Kuhnen, "Inverse control of systems with hysteresis and creep," *IEE Proceedings*, vol. 148, no. 3, pp. 185–192, 2001.
- [14] S. Bashash and N. Jalili, "Robust multiple frequency trajectory tracking control of piezoelectrically driven micro/nano-positioning systems," *IEEE Transactions on Control Systems Technology*, vol. 15, no. 5, pp. 867–878, 2007.
- [15] H. Jiang, H. Ji, J. Qiu, and Y. Chen, "A modified prandtl-ishlinskii model for modeling asymmetric hysteresis of piezoelectric actuators," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 57, no. 5, pp. 1200–1210, 2010.
- [16] M. Al Janaideh, Y. Feng, S. Rakheja, C. Y. Su, and C. A. Rabbath, "Hysteresis compensation for smart actuators using inverse generalized prandtl-ishlinskii model," in *2009 American Control Conference, ACC 2009*, pp. 307–312, usa, June 2009.
- [17] M. Al Janaideh, S. Rakheja, and C. Y. Su, "An analytical generalized Prandtl-Ishlinskii model inversion for hysteresis compensation in micropositioning control," *IEEE/ASME Transactions on Mechatronics*, vol. 16, no. 4, pp. 734–744, 2011.
- [18] Q. Xu and P.-K. Wong, "Hysteresis modeling and compensation of a piezostage using least squares support vector machines," *Mechatronics*, vol. 21, no. 7, pp. 1239–1251, 2011.
- [19] P.-K. Wong, Q. Xu, C.-M. Vong, and H.-C. Wong, "Rate-dependent hysteresis modeling and control of a piezostage using online support vector machine and relevance vector machine," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 4, pp. 1988–2001, 2012.
- [20] J.-J. Tzen, S. L. Jeng, and W. H. Chieng, "Modeling of piezoelectric actuator for compensation and controller design," *Precision Engineering*, vol. 27, no. 1, pp. 70–86, 2003.
- [21] L. Sun, C. Ru, W. Rong, L. Chen, and M. Kong, "Tracking control of piezoelectric actuator based on a new mathematical model," *Journal of Micromechanics and Microengineering*, vol. 14, no. 11, pp. 1439–1444, 2004.
- [22] S. Bashash and N. Jalili, "Underlying memory-dominant nature of hysteresis in piezoelectric materials," *Journal of Applied Physics*, vol. 100, no. 1, Article ID 014103, pp. 1–6, 2006.
- [23] S. Bashash and N. Jalili, "A polynomial-based linear mapping strategy for feedforward compensation of hysteresis in piezoelectric actuators," *Journal of Dynamic Systems, Measurement and Control*, vol. 130, no. 3, Article ID 031008, pp. 1–10, 2008.
- [24] S. Bashash, N. Jalili, P. Evans, and M. J. Dapino, "Recursive memory-based hysteresis modeling for solid-state smart actuators," *Journal of Intelligent Material Systems and Structures*, vol. 20, no. 18, pp. 2161–2171, 2009.

Research Article

Model Predictive Engine Air-Ratio Control Using Online Sequential Relevance Vector Machine

Hang-cheong Wong,¹ Pak-kin Wong,¹ and Chi-man Vong²

¹ Department of Electromechanical Engineering, Faculty of Science and Technology, University of Macau, Taipa 999078, Macau

² Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Taipa 999078, Macau

Correspondence should be addressed to Hang-cheong Wong, hcwong@umac.mo

Received 17 January 2012; Accepted 24 March 2012

Academic Editor: Qingsong Xu

Copyright © 2012 Hang-cheong Wong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Engine power, brake-specific fuel consumption, and emissions relate closely to air ratio (i.e., lambda) among all the engine variables. An accurate and adaptive model for lambda prediction is essential to effective lambda control for long term. This paper utilizes an emerging technique, relevance vector machine (RVM), to build a reliable time-dependent lambda model which can be continually updated whenever a sample is added to, or removed from, the estimated lambda model. The paper also presents a new model predictive control (MPC) algorithm for air-ratio regulation based on RVM. This study shows that the accuracy, training, and updating time of the RVM model are superior to the latest modelling methods, such as diagonal recurrent neural network (DRNN) and decremental least-squares support vector machine (DLSSVM). Moreover, the control algorithm has been implemented on a real car to test. Experimental results reveal that the control performance of the proposed relevance vector machine model predictive controller (RVMMPC) is also superior to DRNNMPC, support vector machine-based MPC, and conventional proportional-integral (PI) controller in production cars. Therefore, the proposed RVMMPC is a promising scheme to replace conventional PI controller for engine air-ratio control.

1. Introduction

Engine power, brake-specific fuel consumption, and emissions relate closely to air ratio among all the engine variables [1]. Air ratio is also called lambda. It indicates the amount that the actual available air-fuel ratio mixture differs from the stoichiometric air-fuel ratio of the fuel being used. Manzie et al. [2, 3] mentioned that if the air-fuel ratio is 1% lower than its stoichiometric ratio (e.g., 14.7:1 for gasoline), carbon monoxide (CO) and hydrocarbon (HC) emissions will be significantly increased. An air-fuel ratio that is 1% higher than the stoichiometric ratio produces more nitrogen oxides (NO_x), up to 50%. Modern automotive engines are controlled by the electronic control unit (ECU) which usually uses look-up tables with compensation of a proportional-integral (PI) closed-loop controller for lambda regulation. Since the nature of engine combustion is multivariable, time-varying, time-delay, and chaotic, look-up tables with PI

controller cannot produce desirable and accurate lambda control [2, 3].

So far, there are only a few papers focusing on air-ratio control, but some control strategies were developed for air-fuel ratio (AFR) control in the past decade, including the sliding mode control [4], radial basis function neural-network feed-forward feedback control [5], and model predictive control (MPC) using neural network-based models [6–8]. In the aforementioned researches, the most appropriate and the latest technique is MPC based on diagonal recurrent neural network (DRNN) [8] because of its fast computational time. The MPC is very robust and suitable for a multivariable, time-varying, and delay system that matches the characteristic of modern engine AFR control systems [9]. A reliable engine performance model is a core component of the MPC. However, the engine models developed in [5–8] were surrogate models, which were trained from the data generated by empirical equations. Moreover, there were

many assumptions in the empirical equations. In fact, many coefficients in the empirical equations are also difficult to determine for a real engine [10]. Therefore the neural-network prediction models derived from the data generated by empirical equations cannot reflect the actual performance of the controller in real engines. Meanwhile, their control target and engine models only focused on AFR instead of air ratio. AFR control is ineffective for engine performance control because it does not consider the fuel variation whereas air-ratio is a fuel independent index. Furthermore, in [5–8], only simulation tests were presented and the testing on real car engines has never been done. Besides, the inherent drawbacks of the neural network (NN) would make itself difficult put into practice including multiple local minima, user burden on selection of optimal NN structure, and overfitting. Another practical challenge to the modelling of engine lambda performance is that the lambda model is required to be updated for any changes in engine performance such as the engine aging or fair user modification on it. Therefore, the current research objective is to develop a reliable and nonlinear time-series prediction model for chaotic engine lambda behaviour and the model should have ability of online update as well.

Least-squares support vector machine (LS-SVM) is an alternative technique of nonlinear modelling [11, 12], which combines the advantages of NNs (handling large amount of highly nonlinear data) and nonlinear regression (high generalization). In recent years, LS-SVM has been successfully applied to a wide range of engineering applications [10, 13–17]. Hence it is believed that LS-SVM can also be applied to estimate the lambda model. However, LS-SVM is just an offline algorithm which cannot continually update the lambda prediction models with the subsequent samples for correction. Although there is an online version of LS-SVM [18] based on pruning, namely, incremental least-squares support vector machine (DLSSVM), it still suffers from two drawbacks. The first one is that there are two user-defined hyperparameters in DLSSVM which seriously affect the model accuracy and generalization. The optimization of these two hyperparameters is time-consuming and the optimality of these hyperparameter values is not guaranteed after incremental update. Since the number of support vectors typically grows linearly with the size of the training dataset, DLSSVM makes unnecessarily liberal use of basis functions which results in long computational time for real-time control applications.

To overcome the deficiencies of neural networks and LS-SVM, Tipping and Faul [19] proposed an advanced modelling technique, namely, relevance vector machine (RVM). RVM is an online machine learning technique which utilizes more flexible and sparser model without setting additional regularization parameters. RVM takes shorter training time than NN. In addition, the model updating time of RVM is also typically fast so that the RVM model can be continually updated whenever new lambda samples arrive.

In view of the deficiencies of the existing work and the chaotic nature of engine combustion [20] as well as the advantages of RVM, a promising avenue of research is to apply relevance vector machine and model predictive

controller together to complicated engine systems and test it on a real car engine for practical examination of the MPC. This paper presents a nonlinear MPC algorithm for air-ratio control based on an RVM model for a real car engine. The method is new for real-time engine lambda control. The RVM model can be updated continually to model the engine performance variation and severely nonlinear and chaotic natures. Based on the multiple-step-ahead prediction of the lambda, an optimal control signal is obtained to regulate the lambda to the desired value upon the change of engine operating condition. In order to show the advantages of the presented control algorithm over the existing algorithms, the control performance of the proposed relevance vector machine model predictive controller (RVMMPC) is compared with the latest neural network-based method for air-fuel ratio control, diagonal recurrent neural network MPC (DRNNMPC) [8] and conventional PI controller in production cars. Besides, it is interesting in extending the incremental least-squares support vector machine (DLSSVM) [18] to MPC and making comparison with the RVMMPC for engine air-ratio. To the best knowledge of the authors, this research is the first attempt at extending online LS-SVM and RVM to the domain of automotive engine air-ratio modelling and control.

2. Relevance Vector Machine

In the first part of this section, the algorithm of RVM modelling is briefly reviewed. In order to handle the online update of RVM model, an online sequential algorithm is adopted and described in the last part of this section.

2.1. RVM Modelling. Given a training dataset \mathbf{D} of N input vectors \mathbf{x}_n , $n = 1$ to N , along with N corresponding scalar-valued output y_n . The input vector $\mathbf{x}_n \in \mathbb{R}^m$ contains the previous measured engine time-series parameters including fuel injection time, throttle position, and air ratio at a specific time instant. The corresponding air ratio at that time is defined as the output $y_n \in \mathbb{R}$ which is assumed to contain zero-mean Gaussian noise with variance σ^2 . Hence, the probability of prediction error ε_n for y_n is a Gaussian distribution of zero mean and variance σ^2 , that is, $p(\varepsilon_n | \sigma^2) = N(0, \sigma^2)$, with

$$y_n = f(\mathbf{x}_n, \mathbf{w}) + \varepsilon_n, \quad (1)$$

$$p(y_n | \mathbf{x}_n, \mathbf{w}, \sigma^2) = N(\hat{y}_n, \sigma^2), \quad (2)$$

where $\hat{y}_n = f(\mathbf{x}_n, \mathbf{w})$ is the output prediction of the true value y_n , and $\mathbf{w} = [w_0, \dots, w_N]$ is the weight vector for the RVM model.

Generally, the prediction can be represented by

$$\begin{aligned} \hat{y} &= f(\mathbf{x}, \mathbf{w}) \\ &= w_0 + \sum_{n=1}^N w_n K(\mathbf{x}, \mathbf{x}_n) \\ &= \mathbf{w} \boldsymbol{\phi}(\mathbf{x}), \end{aligned} \quad (3)$$

where $K(\mathbf{x}, \mathbf{x}_i)$ is a basis function and $\phi(\mathbf{x}) = [1K(\mathbf{x}, \mathbf{x}_1) \cdots K(\mathbf{x}_n, \mathbf{x}_N)]^T$. In the current research, radial basis function (RBF) is chosen as the basis function K because it is commonly used for modelling problem [10].

Under the aforementioned formulation, the likelihood function of the output vector $\mathbf{y} = y_n, n = 1$ to N , is constructed by a multivariate Gaussian:

$$p(\mathbf{y} | \mathbf{w}, \sigma^2) = (2\pi)^{-N/2} \sigma^{-N} \exp\left\{-\frac{\|\mathbf{y} - \hat{\mathbf{y}}\|^2}{2\sigma^2}\right\}, \quad (4)$$

where $\hat{\mathbf{y}} = \Phi \mathbf{w}^T$ and Φ represents an $N \times (N + 1)$ design matrix with the notation $\Phi = [\phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_N)]^T$, and $\phi(\mathbf{x}_n) = [1K(\mathbf{x}_n, \mathbf{x}_1) \cdots K(\mathbf{x}_n, \mathbf{x}_N)]^T, n = 1$ to N .

The likelihood function, (4), is complemented by a prior over the weights, $\mathbf{w} = \{w_p\}, p = 0$ to N , to control the complexity of the model and avoid overfitting. The prior is a zero-mean Gaussian distribution and is defined over every weight w_n as follows:

$$p(\mathbf{w} | \boldsymbol{\alpha}) = (2\pi)^{-N/2} \prod_{p=0}^N \alpha_p^{1/2} \exp\left(-\frac{\alpha_p \mathbf{w}_p^2}{2}\right). \quad (5)$$

The hyperparameter vector $\boldsymbol{\alpha} = [\alpha_0 \cdots \alpha_N]^T$ that controls how far each weight w_p is allowed to deviate from zero. Given $\boldsymbol{\alpha}$, using Bayes' rule, the posterior over \mathbf{w} is expressed as

$$p(\mathbf{w} | \mathbf{y}, \boldsymbol{\alpha}, \sigma^2) = \frac{p(\mathbf{y} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha})}{p(\mathbf{y} | \boldsymbol{\alpha}, \sigma^2)} = N(\mathbf{w} | \boldsymbol{\mu}, \boldsymbol{\Sigma}). \quad (6)$$

The posterior mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ for \mathbf{w} are given as follows [19]:

$$\boldsymbol{\Sigma} = (\mathbf{A} + \sigma^{-2} \Phi^T \Phi)^{-1}, \quad (7)$$

$$\boldsymbol{\mu} = \mathbf{w} = \sigma^{-2} \boldsymbol{\Sigma} \Phi^T \mathbf{y},$$

where \mathbf{A} is defined as $\text{diag}(\alpha_0 \cdots \alpha_N)$.

The posterior mean $\boldsymbol{\mu}$ is an estimation of the weight vector \mathbf{w} for prediction. The only unknowns in (7) are the hyperparameters $\boldsymbol{\alpha}$ which can be estimated via a type-II maximum likelihood procedure [19]. It is called sparse Bayesian learning which is formulated as the local maximization with respect to $\boldsymbol{\alpha}$ of the marginal likelihood or, equivalently, its logarithm $L(\boldsymbol{\alpha})$:

$$L(\boldsymbol{\alpha}) = \log p(\mathbf{y} | \boldsymbol{\alpha}, \sigma^2) = \log \int_{-\infty}^{\infty} p(\mathbf{y} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha}) d\mathbf{w}$$

$$= -\frac{1}{2} [N \log 2\pi + \log |\mathbf{C}| + \mathbf{y}^T \mathbf{C}^{-1} \mathbf{y}], \quad (8)$$

where $\mathbf{C} = \sigma^2 \mathbf{I} + \Phi \mathbf{A} \Phi^T$.

By maximizing (8) over $\boldsymbol{\alpha}$, the most probable values $\boldsymbol{\alpha}_{\text{MP}}$ can be generated. Then, \mathbf{A} can be obtained by substituting $\boldsymbol{\alpha}_{\text{MP}}$ into (7). Afterwards, the covariance $\boldsymbol{\Sigma}$ and posterior

mean $\boldsymbol{\mu} = [\mu_0, \mu_1, \dots, \mu_N]^T$ can be estimated. Therefore, the RVM model f can be found by setting $\mathbf{w} = \boldsymbol{\mu}$:

$$\hat{y} = y_p = f(\mathbf{x}, \mathbf{w}) = f(\mathbf{x}, \boldsymbol{\mu}),$$

$$y_p = \mu_0 + \sum_{n=1}^N \mu_n K(\mathbf{x}, \mathbf{x}_n), \quad (9)$$

$$y_p = \mu_0 + \sum_{n=1}^N \mu_n \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_n\|^2}{\sigma^2}\right),$$

where \hat{y} is the prediction of the output lambda with the unseen input data \mathbf{x} containing the recent measured engine data series, which is explained in Section 4.1. One crucial observation is that the optimal values of many hyperparameters α_p are typically infinite. Considering that $w_p = \mu_p \propto (1/\alpha_p)$, this leads to a parameter posterior infinitely peaked at zero for many weights w_p . In consequence, the posterior mean $\boldsymbol{\mu}$ consists of very few nonzero elements. This results in good sparseness for RVM whereas LS-SVM and NN do not have this advantage. A good sparseness implies that the computational time for prediction can be shortened.

2.2. Online RVM Modelling. In order to train and update the RVM model continually, an online algorithm for RVM [19] is employed. The algorithm starts with an empty set of basis function for the RVM model and sequentially adds basis functions $\phi_n = [1K(\mathbf{x}_n, \mathbf{x}_1) \cdots K(\mathbf{x}_n, \mathbf{x}_N)]^T$ to increase the marginal likelihood and modify their corresponding weights w_n . Within the same principal framework, the likelihood can also be updated by deleting those basis functions which subsequently become redundant. Removing a basis function ϕ_n implies that the corresponding \mathbf{x}_n is no longer important and can be excluded from the design matrix Φ in constructing the RVM model f , which can be simply done by setting the corresponding hyperparameter α_p equal to infinity (practically a very large value). In this way, the corresponding weight value $w_p = \mu_p \propto (1/\alpha_p)$ becomes zero. Since w_0 is a bias corresponding to no training data, only the weights w_1, w_2, \dots, w_N are considered. The data \mathbf{x}_n with nonzero weights $w_n, n = 1$ to N , are referred to relevance vectors. Since the basis functions are sequentially added to or deleted from the lambda model by RVM, the likelihood can be continually updated and hence this mechanism makes online update of the lambda model feasible.

In the following discussions, the constraints to add and delete basis function from the RVM lambda model are given, and then the detailed online RVM algorithm is presented.

2.2.1. Constraints for the Basis Functions Update. In (8), the covariance matrix \mathbf{C} includes all basis functions in the RVM prediction model. When it is necessary to remove a basis function ϕ_n from \mathbf{C} , the new covariance matrix \mathbf{C}_{-n} with the influence of removed basis function ϕ_n can be expressed as follows:

$$\mathbf{C}_{-n}(\boldsymbol{\alpha}) = \sigma^2 \mathbf{I} + \sum_{j \neq n} \alpha_j^{-1} \phi_j \phi_j^T. \quad (10)$$

It was shown in [21] that $L(\boldsymbol{\alpha})$ has a unique maximum with respect to $\boldsymbol{\alpha}$:

$$\alpha_n = \frac{S_n^2}{Q_n^2 - S_n}, \quad \text{if } Q_n^2 > S_n, \quad (11)$$

$$\alpha_n = \infty, \quad \text{if } Q_n^2 \leq S_n, \quad (12)$$

where

$$S_n = \phi_n^T \mathbf{C}_{-n}^{-1} \phi_n, \quad (13)$$

$$Q_n = \phi_n^T \mathbf{C}_{-n}^{-1} \mathbf{y}, \quad (14)$$

with $\phi_n = [1K(\mathbf{x}_n, \mathbf{x}_1) \cdots K(\mathbf{x}_n, \mathbf{x}_N)]^T$, $n = 1$ to N .

The results of (11) and (12) imply that if ϕ_n is currently included in the lambda model (i.e., $\alpha_n < \infty$) and $Q_n^2 \leq S_n$, then ϕ_n can be deleted by setting α_n to ∞ . On the other hand, if ϕ_n is currently excluded from the lambda model (i.e., $\alpha_n = \infty$) and $Q_n^2 > S_n$, then ϕ_n can be added by setting α_n to some optimal finite values. With these constraints, an online RVM algorithm can be implemented here in after.

2.2.2. Online Sequential Algorithm of RVM. The steps for the implementation of the online RVM algorithm are described as follows.

- (1) Initialize σ^2 to some sensible values (e.g., $\text{var}[\mathbf{y}] \times 0.1$) and all α_n are notionally set to infinity.
- (2) Initialize S_n and Q_n with a single basis function ϕ_n from (13) and (14) and compute new α_n from (11) which can be simplified as follows.

$$\alpha_n = \frac{\|\phi_n\|^2}{\|\phi_n^T \mathbf{y}\| / \|\phi_n\|^2 - \sigma^2}. \quad (15)$$

- (3) Explicitly compute $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (which are scalars initially), along with initial values of S_n and Q_n for all N basis functions ϕ_n using (7) and (10)–(14).
- (4) Select a candidate basis function ϕ_n from the set of all N basis functions.
- (5) Compute $\theta_n = Q_n^2 - S_n$.
- (6) If $\theta_n > 0$ and $\alpha_n < \infty$ (i.e., ϕ_n is included in the model), then reestimate α_n using (11).
- (7) If $\theta_n = 0$ and $\alpha_n = \infty$, then add ϕ_n to the model with updated α_n .
- (8) If $\theta_n < 0$ and $\alpha_n < \infty$, then delete ϕ_n from the model and set $\alpha_n = \infty$.
- (9) Estimate the noise level, and update σ^2 as follows:

$$\sigma^2 = \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|^2}{\sum_{n=1}^N \alpha_n \boldsymbol{\Sigma}_{nn}}. \quad (16)$$

- (10) Recompute or update $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$ and all S_n and Q_n using (7) and (10)–(14).

- (11) If converged then terminate, otherwise go to Step 4.

Remarks. For the initialization in Step (2), a potential basis function could be the one with the largest normalized projection onto the output vector which gives the largest initial likelihood. In Step (4), a candidate basis function ϕ_n , both included in and excluded from the current RVM model, must be selected for updating but the selection is purely at random. In Step (10), $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$ and all S_n and Q_n are recomputed in full forms of (7) and (10)–(14). In Step (11), attaining a local maximum of the marginal likelihood must be judged. The online algorithm terminates when the changes in $L(\boldsymbol{\alpha})$ for all basis functions in the RVM model are smaller than 10^{-6} and all other $\theta_n \leq 0$ [19].

The aforementioned online sequential algorithm for RVM ensures to increase the marginal likelihood at each step until a local maximum is attained. Although adding and deleting basis functions appear notionally to be independent, posterior statistics for all basis functions are being maintained concurrently; that is, all elements of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ corresponding to removed basis functions are trivially zero.

3. MPC with RVM Model

The structure of the proposed RVMMPCC is shown in Figure 1. The controller consists of the RVM engine lambda model and the optimizer based on Brent's method [22]. The RVM engine lambda model predicts the engine response over a specified time horizon. The predictions are used by the optimizer to determine the tentative fuel injection time u' , that minimizes the following performance criterion over the specified time horizon, and then the optimal fuel injection time signal u is sent to the engine:

$$\begin{aligned} \min J(u') = & \sum_{j=N_1}^{N_2} (y_r(t+j) - y_p(t+j))^2 \\ & + \rho \sum_{j=1}^{N_u} (u'(t+j-1) - u'(t+j-2))^2, \end{aligned} \quad (17)$$

where N_1 and N_2 define the prediction horizon. t is the time step. N_u is the control horizon. ρ is a user-defined control weighting factor which penalizes excessive movement of the control signal (i.e., the fuel injection time). The variables $u'(t+j-1)$ and $u'(t+j-2)$ in the second part of (17) are the tentative fuel injection time at the time step $t+j-1$ and $t+j-2$, respectively. The second part of (17) ensures the stability of the controller output. $y_r(t+j)$ is the target lambda at the time step $t+j$, and y_p is the predicted lambda by the RVM model which is computed from (9) at the time step $t+j$, from which the input vector \mathbf{x} consists of three time series of engine control parameters: $u'(t+j-1)$, $u'(t+j-2)$, ..., throttle position $\text{TP}(t+j-1)$, $\text{TP}(t+j-2)$..., and

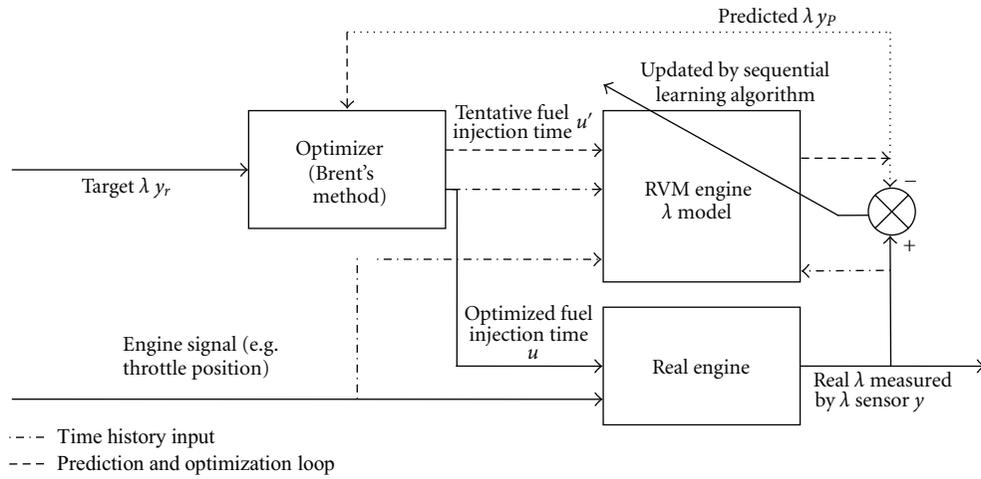


FIGURE 1: Structure of RVMMPCC for engine lambda control.

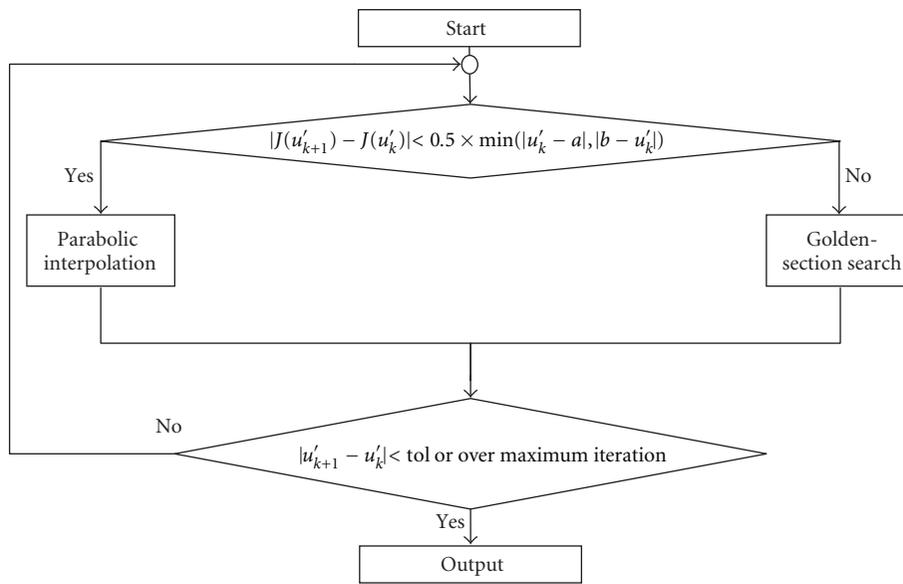


FIGURE 2: General working principle of Brent's method.

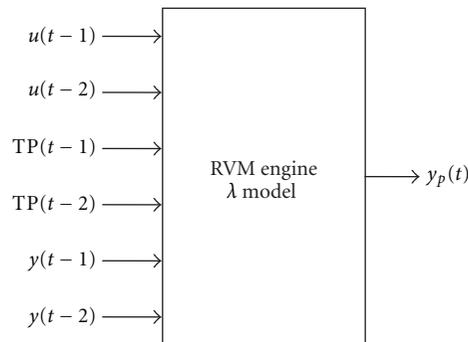


FIGURE 3: Structure of RVM engine lambda model.

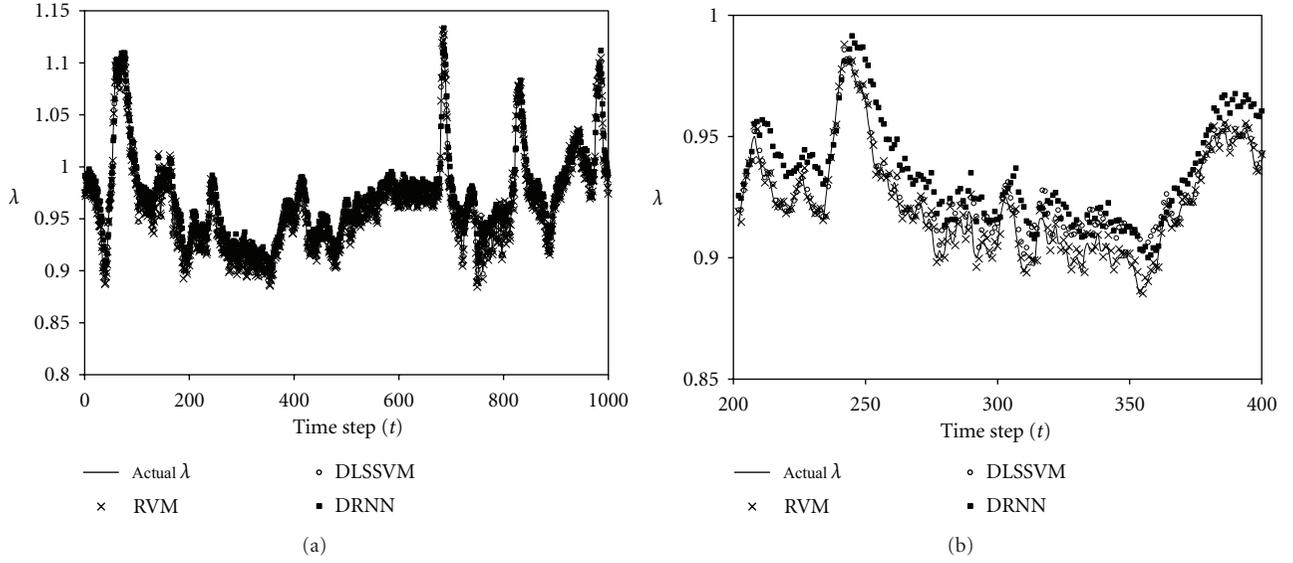


FIGURE 4: (a) Comparison between predicted lambda and the corresponding actual lambda; (b) in detail.

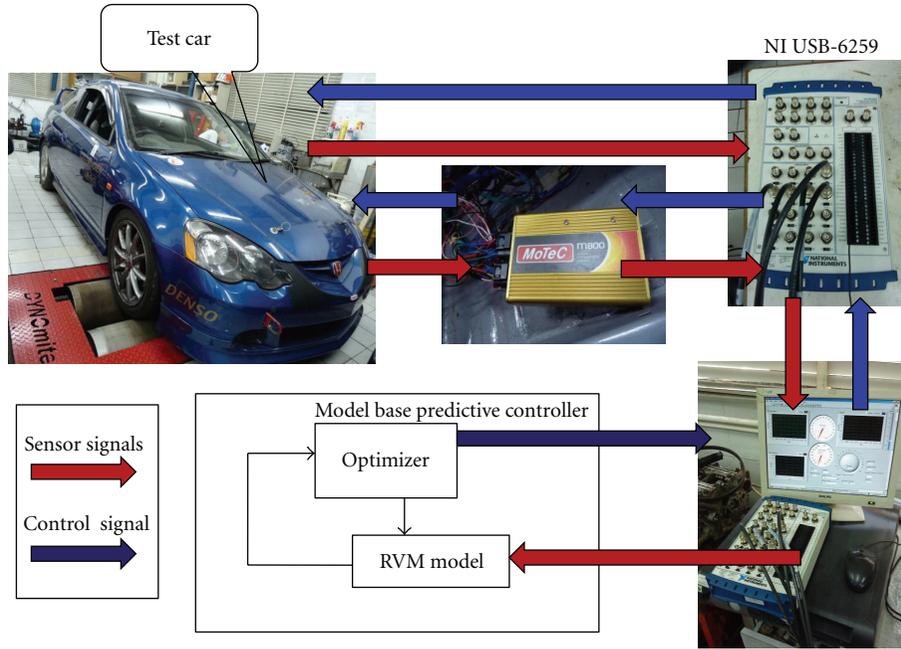


FIGURE 5: Experimental setup.

previous measured lambda $y(t+j-1)$, $y(t+j-2)$ In other words,

$$y_p(t+j) = \mu_0 + \sum_{n=1}^N \mu_n \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_n\|}{\sigma^2}\right). \quad (18)$$

3.1. Single-Dimension Optimization Approach. The original optimization problem involved in this paper is multidimensional and constrained with the tentative control signals over the control horizon N_u and tentative fuel injection times $u'(t)$, $u'(t+1)$, ..., $u'(t+N_u-1)$, which can minimize

the objective function $J(u')$ of (17). Then the predicted lambda values, $y_p(t+N_1)$, $y_p(t+N_1+1)$, ..., $y_p(t+N_2)$, can trace the target lambda values, $y_r(t+N_1)$, $y_r(t+N_1+1)$, ..., $y_r(t+N_2)$, by using the optimized fuel injection time series. Each fuel injection time is normally bounded within the range from 3 ms to 60 ms. However, the multi-dimensional optimization always requires heavy computation, especially when constraints exist. Real-time control applications often put emphasis on computational speed. The research of [7] also showed that the one-dimensional approach is efficient for real time AFR control and the overall

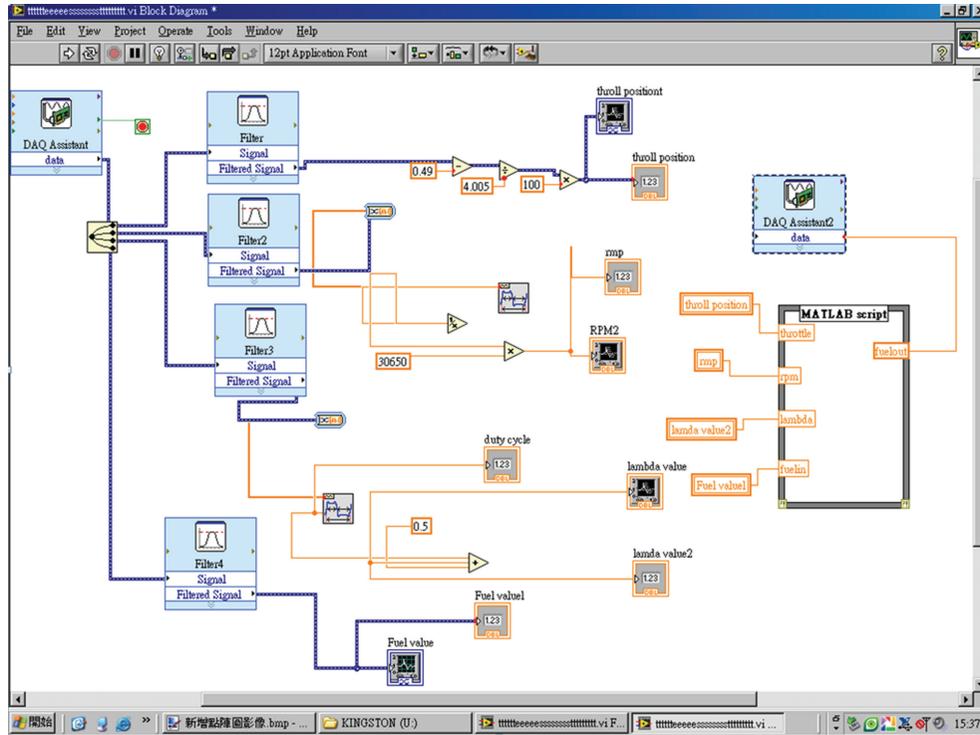


FIGURE 6: LabVIEW interface program.

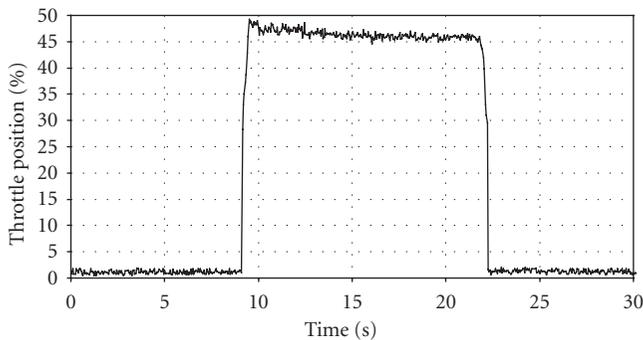


FIGURE 7: Throttle position against time in pilot test 1.

tracking error is similar to that using multidimensional optimization approach. Therefore, the optimization problem to be solved is reduced to onedimensional. In this paper, the control signal u is assumed to remain constant over the control horizon. Therefore, the tentative control signal in the objective function is also constant over the control horizon, that is, $u'(t) = u'(t+1), \dots, = u'(t+N_u - 1)$. In this way, only one parameter $u'(t)$ is needed to determine, and the final fuel injection time at each time step u is set to be the optimal value of $u'(t)$.

3.2. Brent's Method. There are many optimization methods available in the literature; it is impossible to examine all of them. For illustrative purpose, a well-known technique—Brent's method—is examined in this research. Brent's

method is a robust and efficient optimization method. It combines the typical parabolic interpolation and golden-section search. The objective function in each iteration is approximated by an interpolating parabola through three existing points. The minimum point of the parabola is taken as a guess for the minimum point if certain criteria are met. Otherwise, golden-section search is carried out. The advantage of this method is that the high convergence rate of parabolic interpolation can be maintained without losing the robustness of golden-section search [22]. According to this advantage, Brent's method was selected as the MPC optimizer in this study. The general working principle of Brent's method is shown in Figure 2. The detail optimization procedure of Brent's method was presented in [22] and is not presented herein. There are three parameters of Brent's method which are the initial interval of the input variable, $[a, b]$, that is the limit of the fuel injection time, as well as the tolerance, tol , for stopping the optimization procedure. The three variables are set at 3, 60, and 0.05, respectively, because the fuel injection time varies within 3 to 60 ms from 0 to 100% throttle.

4. Implementation and Evaluation of RVM Engine Air-Ratio Model

4.1. RVM Engine Air-Ratio Model Implementation. The objective of the RVM engine lambda model is to predict the future lambda y_p from three time series of inputs: fuel injection time u , throttle position TP, and previous measured lambda y . The structure of the RVM model was chosen to

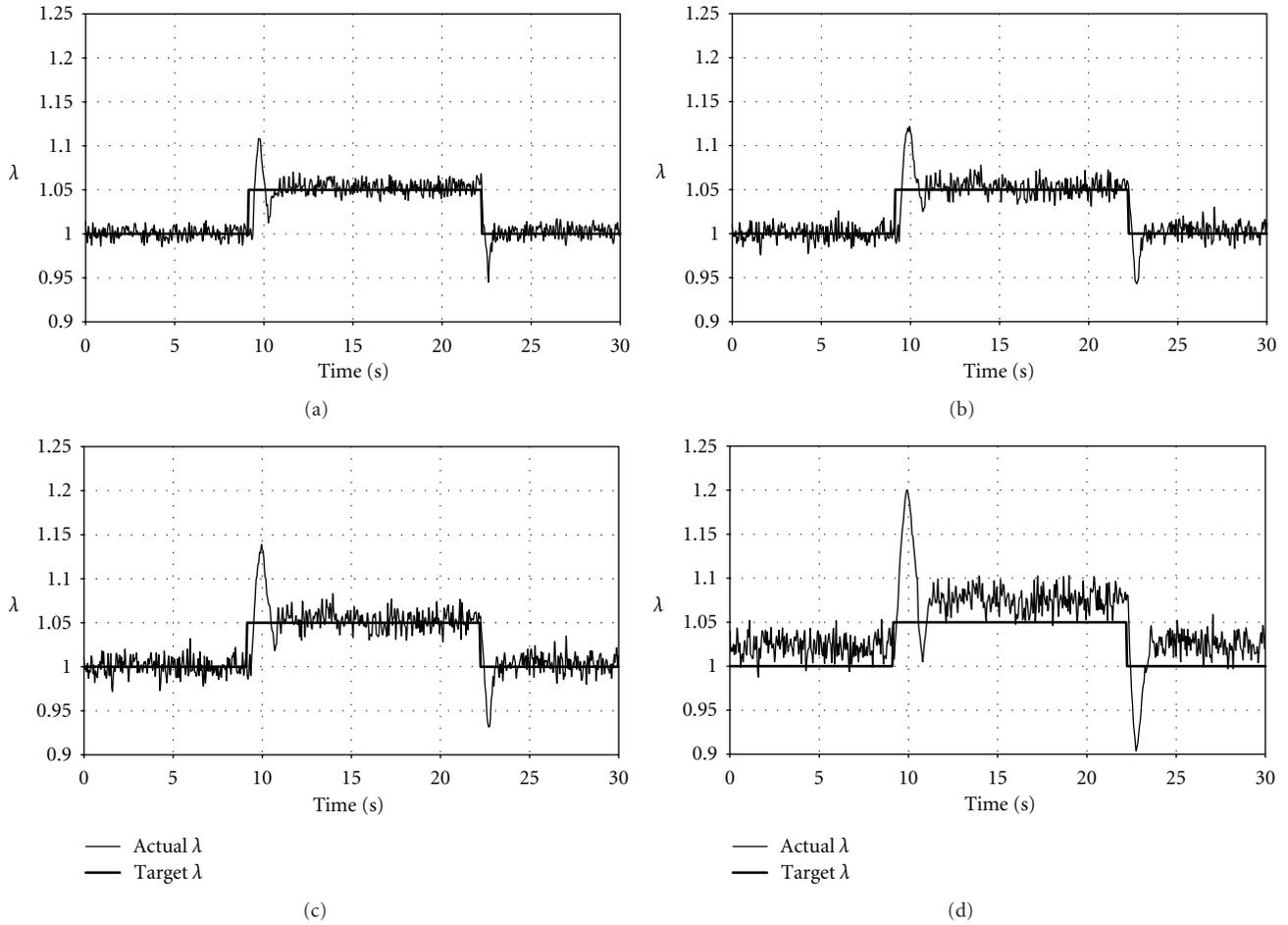


FIGURE 8: Lambda control results of pilot test 1 of (a) RVMMP, (b) DLSSVMMP, (c) DRNNMPC, and (d) PI controller.

be second order (i.e., 2 past time steps) which gives the minimum prediction error [7, 8], and the structure is shown in Figure 3.

To obtain the engine data for building the engine lambda model, 3000 data samples including lambda, fuel injection time, and throttle positions were collected. These 3000 lambda samples were collected over a Honda Type-R K20A i-VTEC engine controlled by a MoTeC M800 programmable electronic control unit with nonfactory calibration data. A dyno test was done for collecting the data samples. In the dyno test, the engine speed was ranged from 1500 r/min to 6500 r/min with random throttle positions bounded between 5% and 65%. The lambda data samples were collected using an in-car wide-band lambda sensor at a sampling frequency of 200 Hz. The first 2000 data samples were used as training dataset \mathbf{D} to build the lambda model. The last 1000 data samples were used as test dataset TEST and the first 900 data samples of TEST were regarded as update dataset UPDATE for real-time updates. The lambda model was updated every 100 measured lambda data samples during prediction. With respect to the training dataset

\mathbf{D} , the test dataset TEST is unseen cases for testing the generalization of the built lambda model.

4.2. Evaluation Criteria for Engine Air-Ratio Model. To illustrate the accuracy, superiority, and online update ability of the proposed RVM model, its prediction result was compared with those obtained from the latest methods, DLSSVM [18] and DRNN [8].

After obtaining the lambda model $RM(\mathbf{x})$ with RVM through the training algorithm presented in Section 2 over \mathbf{D} , $RM(\mathbf{x})$ was then updated for 9 times with UPDATE, 100 data samples each time, to build the updated RVM lambda model $RM^*(\mathbf{x})$. Apart from the RVM model, classical LS-SVM was employed to build the lambda model $DM(\mathbf{x})$ over \mathbf{D} . After that, $DM(\mathbf{x})$ was updated with UPDATE but employing the updating algorithm presented in [18] to build the incremental LS-SVM model $DM^*(\mathbf{x})$. Moreover, DRNN, the same modelling algorithm for AFR model predictive control presented in [8], was applied to estimate the lambda model, $NN(\mathbf{x})$. The network was trained for 60 epochs. After that, $NN(\mathbf{x})$ was updated with UPDATE but

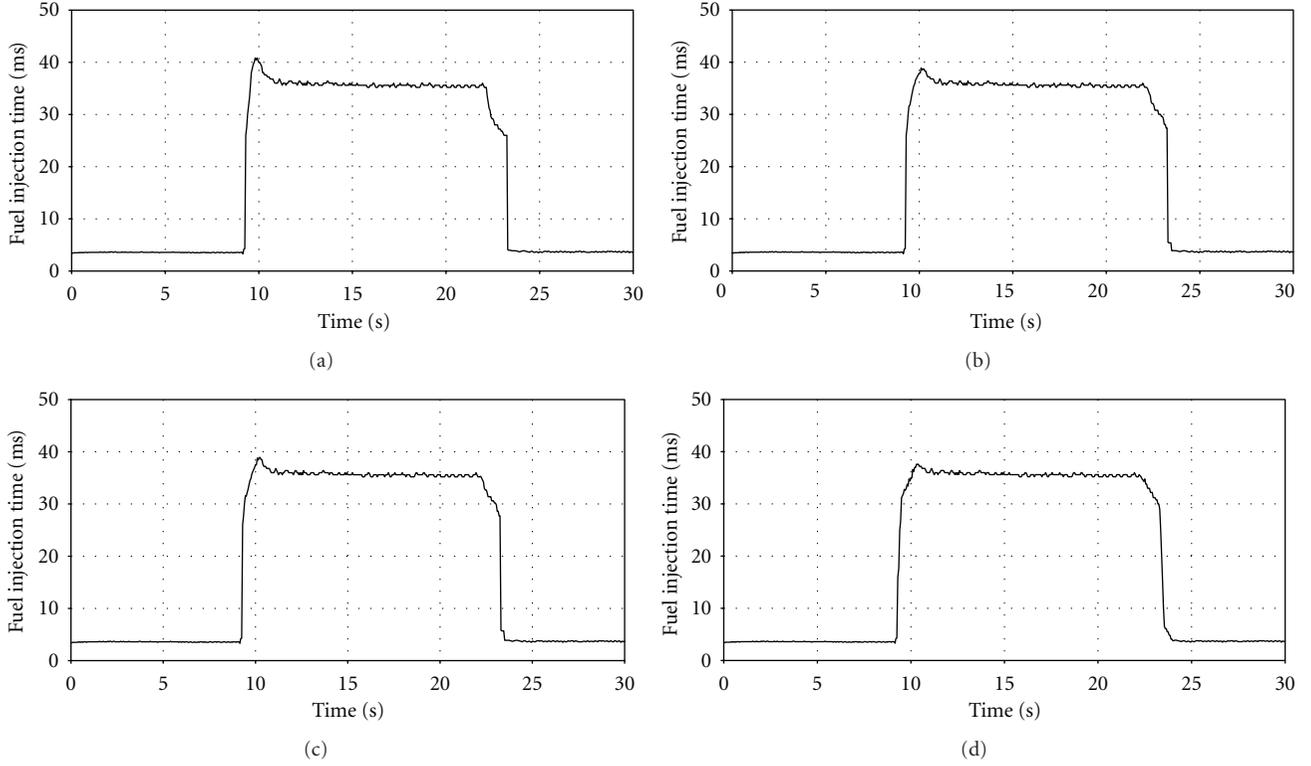


FIGURE 9: Fuel injection time of pilot test 1 of (a) RVMMP, (b) DLSSVMMP, (c) DRNNMPC, and (d) PI controller.

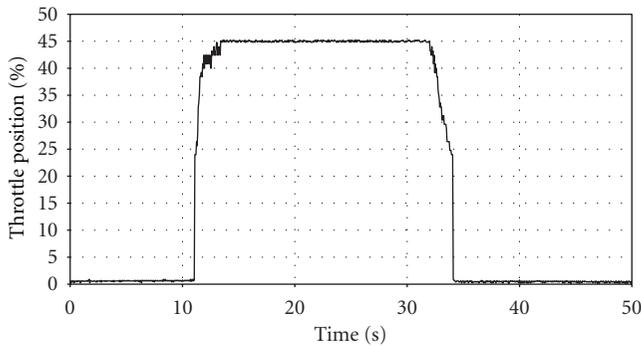


FIGURE 10: Throttle position against time in pilot test 2.

employing the dynamic back-propagation with automatic differentiation technique presented in [8] to build the online model $NN^*(\mathbf{x})$.

After constructing all $RM^*(\mathbf{x})$, $DM^*(\mathbf{x})$, and $NN^*(\mathbf{x})$, the performance of the three lambda models can be evaluated in terms of accuracy. Since the range of lambda for combustible mixture is very narrow, the prediction errors of the above three models are presented by logarithmic mean absolute error ($LMAE$), and they were evaluated one by one against the test dataset, TEST, using (19):

$$LMAE = -\log \left[\frac{1}{T} \sum_{k=1}^T |y_k - f^*(\mathbf{x}_k)| \right], \quad (19)$$

where $f^*(\mathbf{x}_k)$ represents either $RM^*(\mathbf{x}_k)$, $DM^*(\mathbf{x}_k)$, or $NN^*(\mathbf{x}_k)$, \mathbf{x}_k is the k th new input vector for lambda prediction, y_k is the corresponding actual lambda value of $f^*(\mathbf{x}_k)$, and T is the total number of predictions. The value of T is equal to 1000 in this case study. Besides $LMAE$, the accuracies of the three models were also evaluated by the linear regression R^2 value. The prediction results between the predicted lambda values and the corresponding actual lambda values over TEST are shown in Figure 4.

Table 1 shows the $LMAE$ and R^2 values of the three lambda models. According to the $LMAE$ values in Table 1, the RVM lambda model $RM^*(\mathbf{x})$ outperforms $DM^*(\mathbf{x})$ and $NN^*(\mathbf{x})$ by approximately 4% and 29%, respectively. It is believed that the accuracy of the RVM lambda model can be significantly increased if the model update is continually carried out. Moreover, as compared with DLSSVM, RVM has better sparseness. This property is very important for online control and system identification because the computational time and the size of memory are fewer.

In addition to the improvement on prediction accuracy, RVM can also reduce the time for training and updating the lambda model significantly. Table 2 shows the model training time, average model updating time, and cumulative model updating time after nine times of updates. In Table 2, the lambda model training and average model updating time of RVM are 1.52 s and 0.08 s, respectively. The lambda model training time of $RM^*(\mathbf{x})$ is the shortest among the three lambda models. The average model updating time of $RM^*(\mathbf{x})$, 0.08 s, is only about one-sixth of $DM^*(\mathbf{x})$,

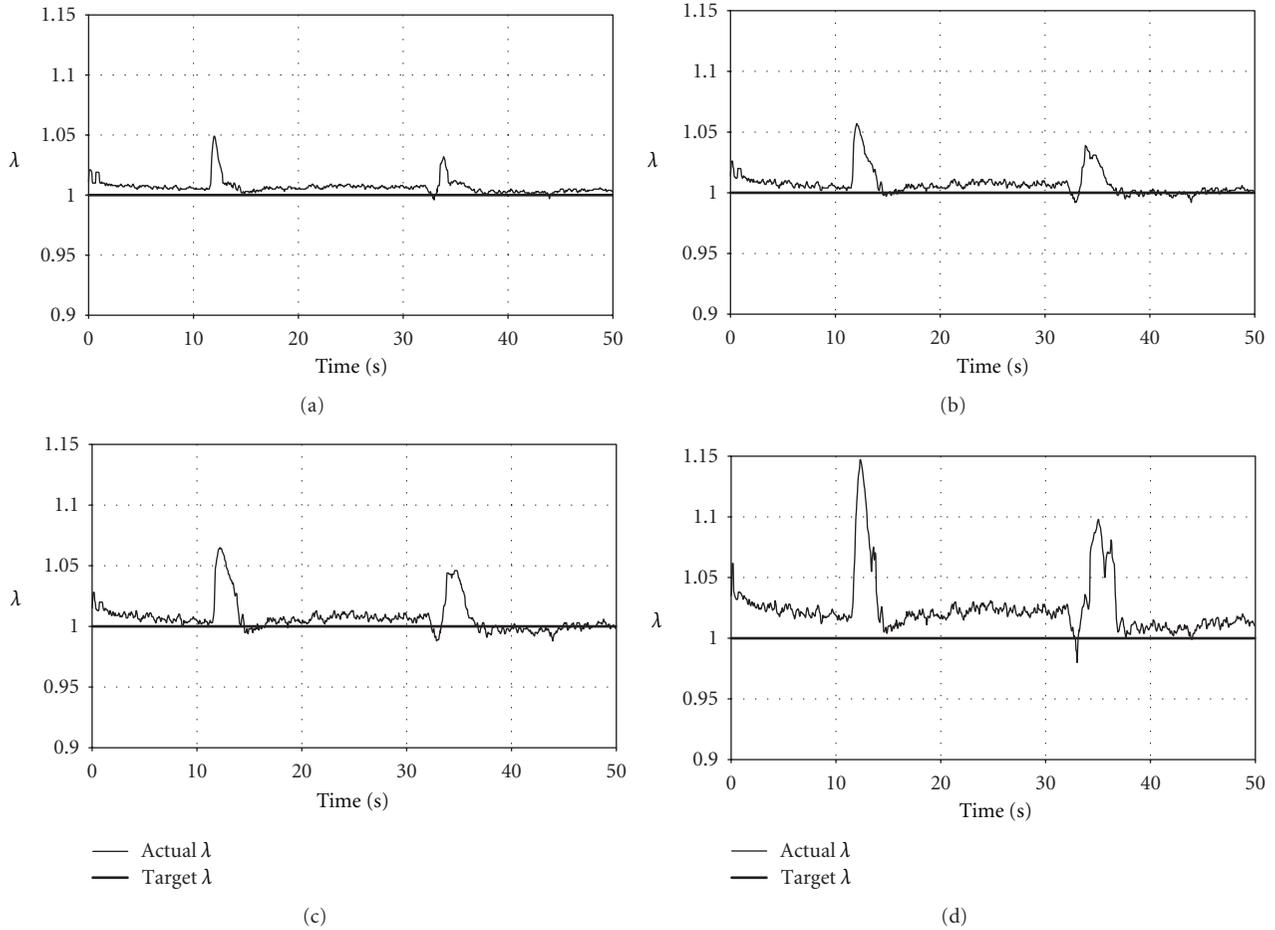


FIGURE 11: Lambda control results of pilot test 2 of (a) RVMMPCC, (b) DLSSVMMPCC, (c) DRNNMPPCC, and (d) PI controller.

0.52 s. It is because in RVM, the updating procedure is applied to every basis function individually while DLSSVM has to work on a large matrix for its estimated inversion. Moreover, the average model updating time of $RM^*(\mathbf{x})$ is also about one-eleventh of $NN^*(\mathbf{x})$. This achievement may be accomplished by the effective online algorithm of RVM. The total cumulative model updating time saved by RVM from DLSSVM and DRNN after 9 update iterations is 3.99 s and 7.56 s, respectively. However, if the lambda model is often updated, the cumulative model updating time saved by RVM will be very significant. Obviously, RVM has more advantage than DLSSVM and DRNN algorithms. As a whole, the high accuracy and short updating time of the lambda model using RVM make online model predictive air-ratio control more feasible.

Experimental results show that the model accuracy, model training time, and updating time of RVM are the best among the aforementioned algorithms. Therefore, RVM was confidently selected to implement the model predictive engine air-ratio controller.

5. Implementation and Evaluation of RVMMPCC

5.1. Experimental Setup. The proposed RVMMPCC algorithm was implemented and tested on a Honda Type-R K20A i-VTEC engine with MoTeC M800 programmable ECU and National Instrument (NI) USB-6259. The model predictive control algorithm was implemented using MATLAB. MoTeC M800 is mainly used for engine control, whereas NI USB-6259 is used for sending control signal to the MoTeC ECU via a LabVIEW interface program according to the MATLAB MPC program embedded. In other words, NI USB-6259 serves as an interface between the MATLAB program and the MoTeC ECU. Apart from fuel injector control, the MoTeC ECU also contains many control maps, such as ignition map and valve timing map, to maintain the engine operation. The experimental setup and the LabVIEW interface program are shown in Figures 5 and 6, respectively.

The initial offline training data for building the RVM model were obtained using a wide-band lambda sensor subject to random throttle positions and are discussed in

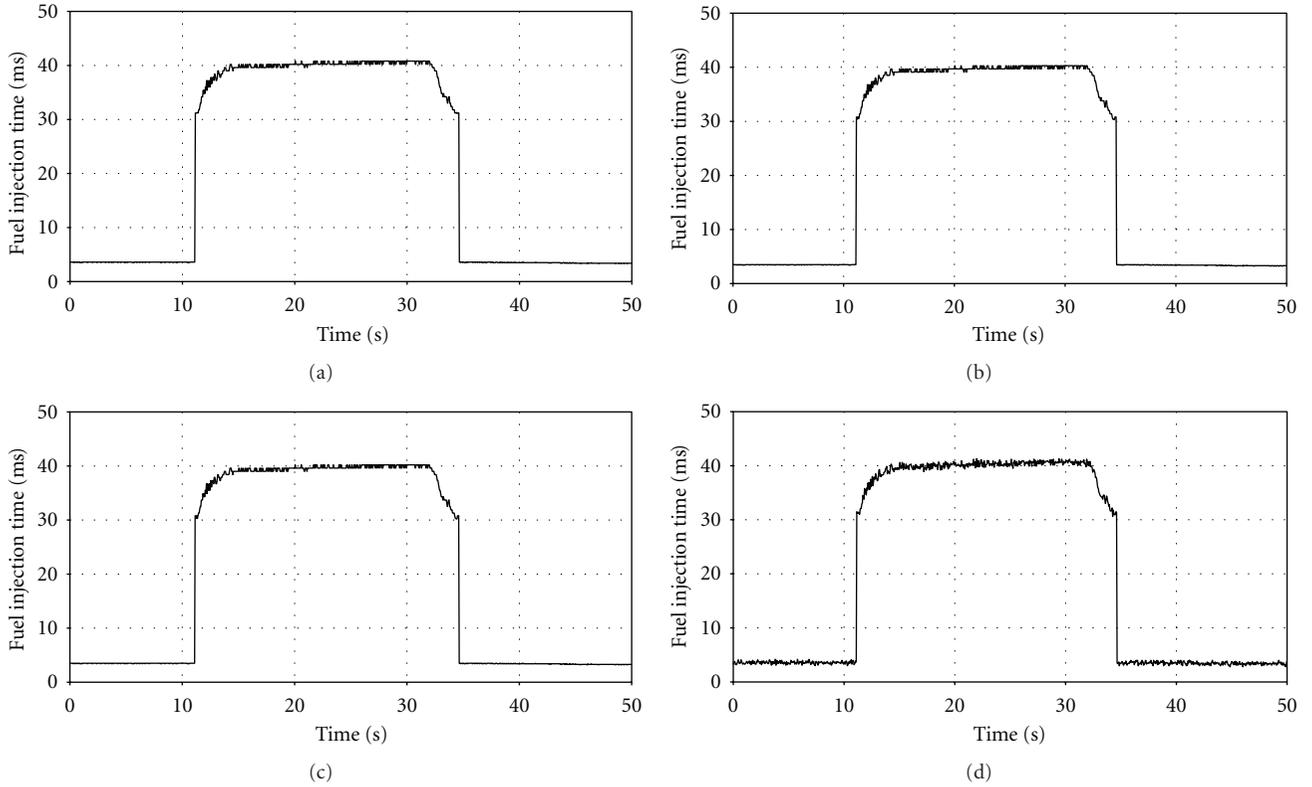


FIGURE 12: Fuel injection time of pilot test 2 of (a) RVMMPCC, (b) DLSSVMMPC, (c) DRNNMPC, and (d) PI controller.

TABLE 1: $LMAE$ and R^2 of different lambda models.

Lambda model	$LMAE$	Improvement in $LMAE$ (%)	R^2	Improvement in R^2 (%)
$RM^*(\mathbf{x})$	2.4112	—	0.9868	—
$DM^*(\mathbf{x})$	2.3166	4.08 ^a	0.9816	0.53 ^a
$NN^*(\mathbf{x})$	1.8674	29.12 ^b	0.9511	3.75 ^b

^aRVM over DLSSVM,

^bRVM over DRNN.

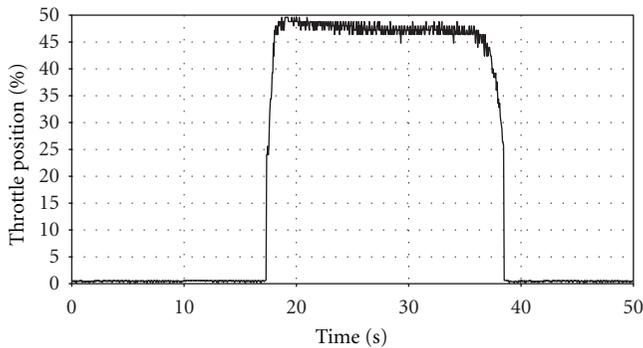


FIGURE 13: Throttle position against time in pilot test 3.

the pervious section. There were three pilot tests in this study. The tests evaluate the tracking error, robustness, and adaptability of the controllers respectively.

5.2. Pilot Test 1: System Tracking Ability. In pilot test 1, the test cycle is shown in Figure 7 where the throttle position changes from 0 to 50% throttle (i.e., partial throttle). In this test, the lambda value is needed to track the target values from the stoichiometric value (1.00) to a value for the best brake-specific fuel consumption (1.05) as the throttle position is changed. This lambda value for such partial throttle is a normal requirement for automobiles. After choosing the sampling time to be 0.005 s, the tracking ability of the RVMMPCC can be examined. By testing many values around the setting used in [8], the parameters of the optimizer were chosen as $N_1 = 1$, $N_2 = 10$, $\rho = 0.5$, and $N_u = 3$. With the test cycle shown in Figure 7 and the parameters chosen, the lambda control result and the corresponding fuel injection time are shown in Figures 8(a) and 9(a), respectively. In order to show the advantage of the RVMMPCC, its control result is compared with other two latest model predictive controllers, decremental LS-SVM model predictive controller (DLSSVMMPC) and DRNN model predictive controller (DRNNMPC), as well as the

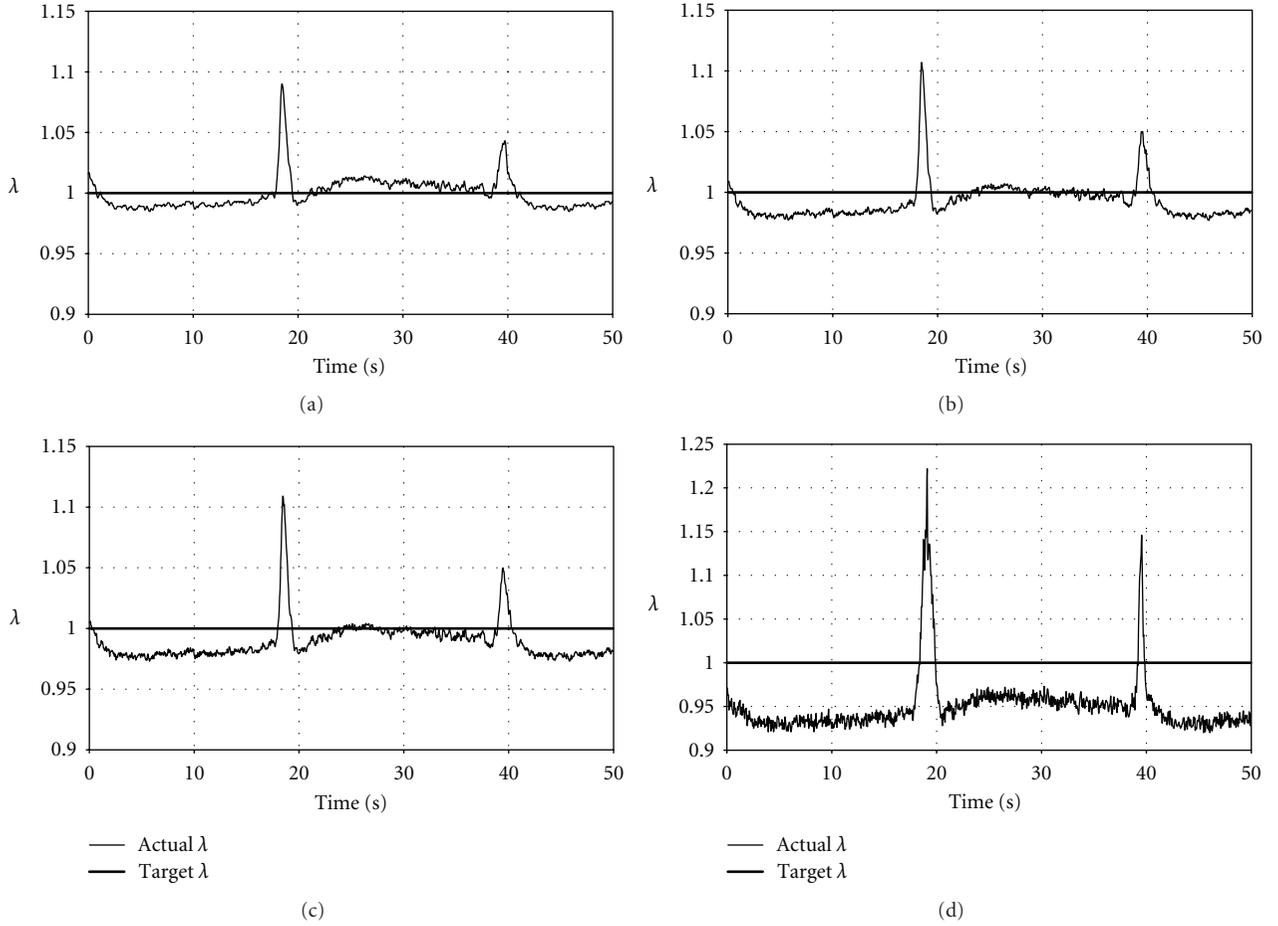


FIGURE 14: Lambda control results of pilot test 3 of (a) RVMMP, (b) DLSSVMMP, (c) DRNNMPC, and (d) PI control.

TABLE 2: Training and updating time of different lambda models.

Lambda model	Training time (s)	Average updating time (s)	Cumulative updating time (s)
$RM^*(\mathbf{x})$	1.52	0.08	0.71
$DM^*(\mathbf{x})$	2.02	0.52	4.70
$NN^*(\mathbf{x})$	3.52	0.92	8.27

PI control algorithm used in the existing automotive ECU. The engine lambda models used in the DLSSVMMP and DRNNMPC are the models mentioned in Section 4.2. The parameters of nonlinear optimization for the DLSSVMMP, and DRNNMPC were the same as those of RVMMP. The PI gains of the PI controller were obtained by the Ziegler-Nichols method. The lambda control results and the corresponding fuel injection time of the DLSSVMMP, DRNNMPC, and PI controller are shown in Figures 8(b)–8(d) and 9(b)–9(d), respectively.

Figure 8(a) shows that the RVMMP can control the lambda to follow the target lambda with the smallest tracking error and overshoot among all the controllers. The control performances of the four controllers are shown in Table 3. As the range of lambda for combustible mixture is very narrow, logarithmic mean absolute error is chosen as the logarithmic

tracking error (LTE) to evaluate the tracking ability of the controllers and is defined by

$$LTE = -\log \left[\frac{1}{T} \sum_{t=1}^T |y_t - y_r(t)| \right], \quad (20)$$

where t is time step, T is the total number of time step, y_t is the actual lambda at each time step, and $y_r(t)$ is the corresponding target lambda at each time step. Table 3 reveals that the LTE of the RVMMP outperforms the DLSSVMMP, DRNNMPC and PI controller by approximately 5%, 8%, and 38%, respectively.

5.3. Pilot Test 2: System Robustness. In pilot test 2, the test cycle is shown in Figure 10 where the throttle position changes from 0 to 50% throttle. This change can be viewed

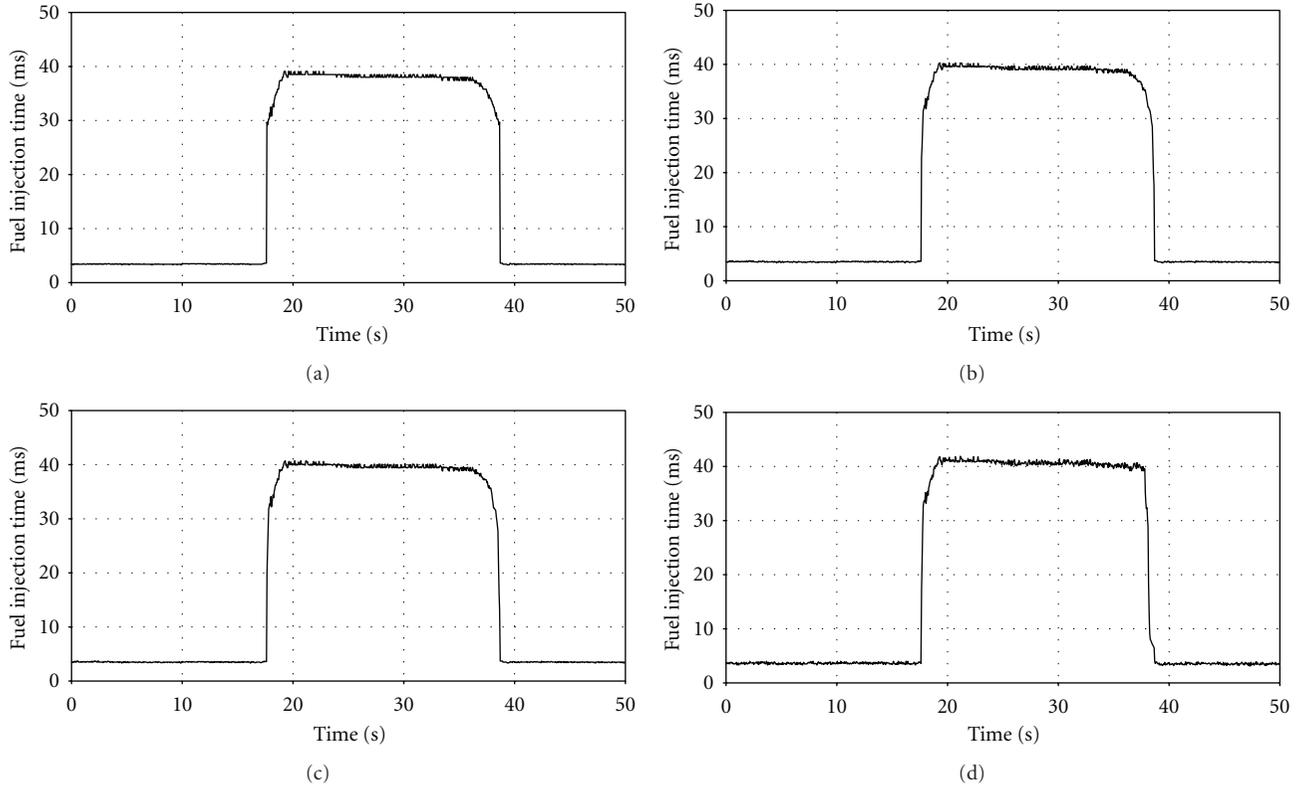


FIGURE 15: Fuel injection time of pilot test 3 of (a) RVMMP, (b) DLSSVMMP, (c) DRNNMPC, and (d) PI controller.

as a disturbance. In this test, the lambda value is needed to control within the $\pm 5\%$ bounds of the stoichiometric value (1.0). After choosing the sampling time to be 0.005 s, the effectiveness of the RVMMP can be examined. The parameters of the optimizer were chosen as $N_1 = 1$, $N_2 = 10$, $\rho = 0.5$, and $N_u = 3$ again. The lambda control result and the fuel injection time of the RVMMP are shown in Figures 11(a) and 12(a), respectively. Similar to pilot test 1, the control result is compared with the two model predictive controllers, DLSSVMMP and DRNNMPC, as well as typical PI controller. The control results and the corresponding fuel injection time of the DLSSVMMP, DRNNMPC, and PI controller are shown in Figures 11(b)–11(d) and 12(b)–12(d), respectively. Figure 11(a) shows that the RVMMP can regulate the lambda to the target lambda with the smallest deviation (i.e., minimum *LTE*) in the steady state among all the controllers. Moreover, the RVMMP can also achieve the smallest overshoot in the transient state among all the controllers. These results show the superior robustness of the RVMMP. The control performances of the four controllers are shown in Table 4, and the control performance of the RVMMP in terms of *LTE* is superior to the DLSSVMMP, DRNNMPC, and PI controller by 3%, 6%, and 35%, respectively.

5.4. Pilot Test 3: Online Update Ability. In the previous two tests, the “bypass air valve” of the test engine is 60% opening. In order to test the update ability of the RVMMP, the

“bypass air valve” was set from 60% to 30%. It is equivalent to the clogging of the engine intake filter as the engine aging. Normally, the lambda value must decrease under the same throttle position as lack of intake air. The testing procedure and the parameter setting are exactly the same as those in pilot test 2 except the change of the “bypass air valve” position. The variation of throttle position in the test cycle is shown in Figure 13. The lambda control result and the fuel injection time of the RVMMP are shown in Figures 14(a) and 15(a), respectively. Similar to pilot tests 1 and 2, the control result is compared with the two model predictive controllers, DLSSVMMP and DRNNMPC, as well as typical PI controller. The control results and the corresponding fuel injection time of the DLSSVMMP, DRNNMPC and PI controller are shown in Figures 14(b)–14(d) and 15(b)–15(d) respectively. Figures 14(a)–14(d) depict the lambda values decrease after changing the position of the “bypass air valve” due to fuel rich. Figures 14(a)–14(c) illustrate that the three controllers, RVMMP, DLSSVMMP, and DRNNMPC, can regulate the lambda with obviously less deviation from the stoichiometric value and overshoot than that of the PI controller because the engine lambda model can be self-updated for any changes in engine condition. The control performances of the four controllers are shown in Table 5, and the control performance of the RVMMP in terms of *LTE* outperforms the DLSSVM, DRNNMPC, and PI controller by approximately 6%, 10%, and 62%, respectively. This promising result indicates that

TABLE 3: Control performance of pilot test 1 of different controllers.

Controller	<i>LTE</i>	Improvement in <i>LTE</i> (%)	Maximum overshoot	Improvement in maximum overshoot (%)
RVMMP	2.1328	—	0.0580	—
DLSSVMMP	2.0317	4.98 ^a	0.0720	19.44 ^a
DRNNMPC	1.9665	8.46 ^b	0.0890	34.83 ^b
PI	1.5485	37.73 ^c	0.1500	61.33 ^c

^a RVMMP over DLSSVMMP.^b RVMMP over DRNNMPC.^c RVMMP over PI.

TABLE 4: Control performance of pilot test 2 of different controllers.

Controller	<i>LTE</i>	Improvement in <i>LTE</i> (%)	Maximum overshoot	Improvement in maximum overshoot (%)
RVMMP	2.1699	—	0.0490	—
DLSSVMMP	2.1158	2.56 ^a	0.0570	14.04 ^a
DRNNMPC	2.0523	5.73 ^b	0.0650	24.62 ^b
PI	1.6092	34.84 ^c	0.1470	66.67 ^c

^a RVMMP over DLSSVMMP.^b RVMMP over DRNNMPC.^c RVMMP over PI.

TABLE 5: Control performance of pilot test 3 of different controllers.

Controller	<i>LTE</i>	Improvement in <i>LTE</i> (%)	Maximum overshoot	Improvement in maximum overshoot (%)
RVMMP	2.0042	—	0.0900	—
DLSSVMMP	1.8965	5.76 ^a	0.1070	15.89 ^a
DRNNMPC	1.8274	9.67 ^b	0.1090	17.43 ^b
PI	1.2386	61.81 ^c	0.2220	59.46 ^c

^a RVMMP over DLSSVMMP.^b RVMMP over DRNNMPC.^c RVMMP over PI.

the RVMMP can regulate air ratio very well even the engine ages and undergoes external disturbance simultaneously.

5.5. Discussion of Results. All the experimental results show that the overall lambda control performance of the RVMMP is better than those of DLSSVMMP, DRNNMPC, and conventional PI controller. There are two important factors affecting the control performance of the model predictive controllers which are the lambda model accuracy, and computational time. As presented in Sections 2.1 and 4.2, the sparseness, model accuracy and model updating time of RVM are better than those of DLSSVM and DRNN. Therefore, the model accuracy, and computational time of the RVMMP is the best among the three controllers. Overall, the RVMMP is the most suitable method for engine air-ratio control. Although Tables 3 to 5 show that the logarithmic tracking errors of the RVMMP have a small improvement only, and the lambda is a very delicate value, a small change can result in a big change of engine performance, so the *LTE* is not very large. For instance, the air

ratio is changed from 1 to 1.05; an increment of 0.05 results in changing from the best emission performance to the best brake-specific fuel consumption. Besides, [2, 3] stated that if AFR is 1% lower than its stoichiometric ratio, CO and HC emissions are significantly increased. An AFR that is 1% higher than the stoichiometric ratio produces more NO_x, up to 50%. As a result, the actual improvement achieved by the RVMMP is very significant.

6. Conclusions

This research is the first attempt at developing RVMMP for engine air-ratio control. This study is also the first to extend DLSSVM to MPC for engine air-ratio control. The RVM is trained by sequential learning algorithm and real engine data, which can let the engine lambda model update continually so as to maintain the lambda model accuracy for any changes in engine performance as the engine aging or fair user modification on it. With the highly accurate engine lambda model, the new MPC strategy can perform

lambda control effectively. The proposed intelligent control algorithm was successfully implemented and tested on a real automotive engine, whereas many previous researches were simulation tests only. Experimental results show that the lambda control performance of the RVMMPC is better than of DLSSVMMPC, DRNNMPC and conventional PI controller. Tables 3 to 5 reveal that the RVMMPC can effectively reduce the lambda deviation and overshoot from target lambda value up to 62% and 67%, respectively. Thus, the RVMMPC is a potential control scheme to replace conventional PI controller in the automotive ECU for engine lambda control. In the future, some advanced optimization algorithms for the RVMMPC and system stability will be studied.

Acknowledgments

This work is supported by the University of Macau Research Grant, Grant no. MYRG149(Y1-L2)-FST11-WPK. The authors would also like to thank the support from Mr. Adrian Kowalski.

References

- [1] W. H. Course and D. L. Anglin, *Automotive Mechanics*, McGraw-Hill, New York, NY, USA, 10th edition, 1993.
- [2] C. Manzie, M. Palaniswami, and H. Watson, "Gaussian networks for fuel injection control," *Journal of Automobile Engineering*, vol. 215, no. D8, pp. 1053–1068, 2001.
- [3] C. Manzie, M. Palaniswami, D. Ralph, H. Watson, and X. Yi, "Model predictive control of a fuel injection system with a radial basis function network observer," *Journal of Dynamic Systems, Measurement and Control*, vol. 124, no. 4, pp. 648–658, 2002.
- [4] S. W. Wang and D. L. Yu, "A new development of internal combustion engine air-fuel ratio control with second-order sliding mode," *Journal of Dynamic Systems, Measurement and Control*, vol. 129, no. 6, pp. 757–766, 2007.
- [5] Y. J. Zhai and D. L. Yu, "Radial-basis-function-based feedforward-feedback control for air-fuel ratio of spark ignition engines," *Proceedings of the Institution of Mechanical Engineers D*, vol. 222, no. 3, pp. 415–428, 2008.
- [6] S. W. Wang, D. L. Yu, J. B. Gomm, G. F. Page, and S. S. Douglas, "Adaptive neural network model based predictive control for air-fuel ratio of SI engines," *Engineering Applications of Artificial Intelligence*, vol. 19, no. 2, pp. 189–200, 2006.
- [7] Y. J. Zhai and D. L. Yu, "Neural network model-based automotive engine air/fuel ratio control and robustness evaluation," *Engineering Applications of Artificial Intelligence*, vol. 22, no. 2, pp. 171–180, 2009.
- [8] Y. J. Zhai, D. W. Yu, H. Y. Guo, and D. L. Yu, "Robust air/fuel ratio control with adaptive DRNN model and AD tuning," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 2, pp. 283–289, 2010.
- [9] G. Y. Li, *Application of Intelligent Control and MATLAB to Electronically Controlled Engines*, Publishing House of Electronics Industry, 2007.
- [10] P. K. Wong, L. M. Tam, K. Li, and C. M. Vong, "Engine idle-speed system modelling and control optimization using artificial intelligence," *Proceedings of the Institution of Mechanical Engineers*, vol. 224, no. D1, pp. 55–72, 2010.
- [11] J. Suykens, J. De Brabanter, L. Lukas, and L. Vandewalle, *Least Squares Support Vector Machines*, World Scientific Press, 1st edition, 2002.
- [12] J. Vallyon and G. A. Horvath, "Sparse least squares support vector machine classifier," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, vol. 1, pp. 543–548, Budapest, Hungary, July 2004.
- [13] C. M. Vong, P. K. Wong, and Y. P. Li, "Prediction of automotive engine power and torque using least squares support vector machines and Bayesian inference," *Engineering Applications of Artificial Intelligence*, vol. 19, no. 3, pp. 277–287, 2006.
- [14] P. K. Wong, C. M. Vong, L. M. Tam, and K. Li, "Data preprocessing and modelling of electronically-controlled automotive engine power performance using kernel principal components analysis and least squares support vector machines," *International Journal of Vehicle Systems Modelling and Testing*, vol. 3, no. 4, pp. 312–330, 2008.
- [15] T. Quan, X. Liu, and Q. Liu, "Weighted least squares support vector machine local region method for nonlinear time series prediction," *Applied Soft Computing Journal*, vol. 10, no. 2, pp. 562–566, 2010.
- [16] X. Wang, H. Zhang, C. Zhang, X. Cai, J. Wang, and J. Wang, "Prediction of chaotic time series using LS-SVM with automatic parameter selection," in *Proceedings of the 6th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT '05)*, pp. 962–965, Dalian, China, December 2005.
- [17] Z. Shi and M. Han, "Support vector echo-state machine for chaotic time-series prediction," *IEEE Transactions on Neural Networks*, vol. 18, no. 2, pp. 359–372, 2007.
- [18] P. K. Wong, H. C. Wong, and C. M. Vong, "Online time-sequence incremental and decremental least squares support vector machines for engine air-ratio prediction," *International Journal of Engine Research*, vol. 13, no. 1, pp. 28–40, 2011.
- [19] M. E. Tipping and A. C. Faul, "Fast marginal likelihood maximisation for sparse bayesian method," in *Proceedings of the 9th International Workshop on Artificial Intelligence and Statistics*, Key West, Fla, USA, January 2003.
- [20] Rototest Research Institute, "Performance measurements on chassis dynamometers," 2005, <http://www.rototest.com>.
- [21] A. C. Faul and M. E. Tipping, "Analysis of sparse Bayesian learning," *Advances in Neural Information Processing Systems*, vol. 14, pp. 383–389, 2002.
- [22] T. R. Chandrupatla, "An efficient quadratic fit—sectioning algorithm for minimization without derivatives," *Computer Methods in Applied Mechanics and Engineering*, vol. 152, no. 1–2, pp. 211–217, 1998.

Research Article

Modelling and Prediction of Particulate Matter, NO_x, and Performance of a Diesel Vehicle Engine under Rare Data Using Relevance Vector Machine

Ka In Wong,¹ Pak Kin Wong,¹ and Chun Shun Cheung²

¹ Department of Electromechanical Engineering, University of Macau, Taipa, Macau

² Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Correspondence should be addressed to Ka In Wong, imkain@gmail.com

Received 13 January 2012; Accepted 27 February 2012

Academic Editor: Qingsong Xu

Copyright © 2012 Ka In Wong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Traditionally, the performance maps and emissions of a diesel engine are obtained empirically through many tests on the dynamometers because no exact mathematical engine model exists. In the current literature, many artificial-neural-network-(ANN-) based approaches have been developed for diesel engine modelling. However, the drawbacks of ANN would make itself difficult to be put into some practices including multiple local minima, user burden on selection of optimal network structure, large training data size, and overfitting risk. To overcome the drawbacks, this paper proposes to apply one emerging technique, relevance vector machine (RVM), to model the diesel engine, and to predict the emissions and engine performance. With RVM, only a few experimental data sets can train the model due to the property of global optimal solution. In this study, the engine speed, load, and coolant temperature are used as the input parameters, while the brake thermal efficiency, brake-specific fuel consumption, concentrations of nitrogen oxides, and particulate matter are used as the output parameters. Experimental results show the model accuracy is fairly good even the training data is scarce. Moreover, the model accuracy is compared with that using typical ANN. Evaluation results also show that RVM is superior to typical ANN approach.

1. Introduction

Air pollution is one of the most challenging problems today in many cities. The increased use of motor vehicles causes the amount of exhaust emissions to increase dramatically, which makes the problem more serious. Reducing the exhaust emissions from engines has then become an important concern of governments and motor vehicle manufacturers. Moreover, in view of the increasing oil price and the need to reduce emission of the global warming gas CO₂, there is a demand to reduce fuel consumption while maintaining the engine performance. Therefore, many researchers have focused on the relations between these two issues, namely, engine performance and emissions.

Diesel engines, though having the advantages of high fuel efficiency and high durability when compared to other engines, are the major source of nitrogen oxides (NO_x) and particulate matter (PM), which are harmful to human health and the environment. In particular, the fine and

ultrafine particles (~10 micrometers or less) emitted by diesel engines can accumulate in the human respiratory system and cause various health problems [1] and influence global climate by absorbing solar radiation and reacting with other atmospheric constituents [2, 3]. Diesel engines are used extensively in buses and trucks; thus they are the major road-side emitters, posing significant threat to the health of the road users. In order to reduce these emissions, the combustion process of the engines has to be controlled. Additional hardware and instruments must be installed to monitor and control the engine operating parameters. Many experiments and tests must also be conducted to obtain a comprehensive understanding on the performance and emissions of the diesel engine. These are very complicated, time consuming, and expensive [4].

A way to solve these problems is to create a mathematical model for the diesel engine so that all the costly and immeasurable data can be predicted and virtual sensors can be used to replace the costly sensors. However, the combustion

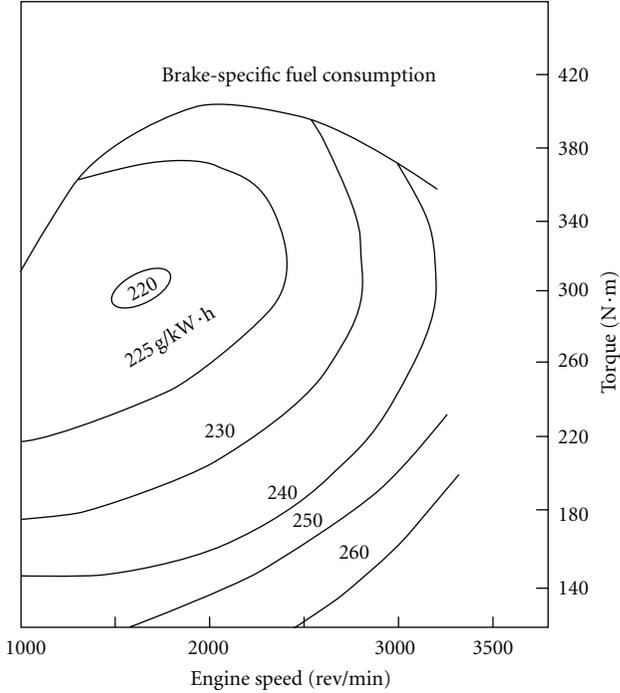


FIGURE 1: An example of diesel engine performance map [5].

process of a diesel engine is too complex that an exact mathematical model still does not exist today. Figure 1 shows one example of diesel engine performance map with only three variables in which the relationship among the engine load (torque), engine speed, and brake specific fuel consumption is already highly nonlinear. It could be imagined that if more variables are studied together, the model will be very complicated and very difficult to obtain. Moreover, the mathematical model varies for different engines.

In general, black-box identification is one of the commonly used modelling techniques suitable for engines because it can manage complex and uncertain information. Many recent researches in black-box identification have described the use of artificial neural network (ANN) for modelling of diesel engine performance [5–9] and emissions [7, 8, 10–13] based on experimental data sets. The ANN has in general, however, three main drawbacks for its learning process [14].

- (1) The architecture, including the number of hidden neurons, has to be determined a priori or modified while training by heuristic, which results in a suboptimal network structure.
- (2) The training process (i.e., the minimization of the residual squared error cost function) in ANN can easily become stuck in local minima. Various ways of preventing local minima, like early stopping, weight decay, have been employed. However, these methods greatly affect the generalization of the estimated function (i.e., the capacity of handling new input cases).
- (3) The amount of training data is usually large. Normally at least 200~400 sets of training data is re-

quired to build an accurate ANN engine model [15]. However, the collection of diesel engine emission and performance data is usually time consuming and costly, so the data set is usually lower than 50, resulting in that ANN may not be a good solution for diesel engine modelling.

To overcome the disadvantages of ANN, an algorithm entitled relevance vector machine (RVM) was proposed by Tipping [16]. This approach is an emerging machine learning technique that is able to utilize more flexible candidate models, which are typically much sparser, offer probabilistic prediction, and avoid the need to set additional hyperparameters. The other advantage is that the training algorithm of RVM can ensure a global optimal solution whereas the learning process of ANN may cause a local optimal solution, so ANN requires more training data to minimize the risk [14]. With this good property, RVM is likely not to require too much sample data to build an accurate model. However, one deficiency of this approach is that the training time is approximately in the cube of the sample numbers. Thankfully, a fast training algorithm [17] is developed for RVM which initializes with an “empty” model, and sequentially “add” samples to increase the marginal likelihood, and also modify their weights. Within the same principal framework, the objective function can also be increased by deleting the samples which subsequently become redundant.

Recently, RVM has been applied to system modelling and predictive control [18–20]. These researches show that RVM is generally superior to the ANN. Moreover, the application of RVM to modelling of diesel engines under rare data is very few. For these reasons, in the present paper, RVM is employed to model the performance and emission characteristics of NO_x and PM of the diesel engine. Experiments are still required to provide sample data for RVM training. To demonstrate the effectiveness of this approach, a neural-network-based diesel engine model is also constructed and compared with the RVM model.

2. Relevance Vector Machine

The procedure of the RVM modelling is introduced here. Consider a training data set \mathbf{D} of N input vectors $\{\mathbf{X}_n\}_{n=1}^N$, along with N corresponding scalar-valued output $\{y_n\}_{n=1}^N$. The output y_n is assumed to contain zero-mean Gaussian noise with variance σ^2 . Hence, the probability of prediction error ε_n for y_n is a Gaussian distribution of zero mean and variance σ^2 : $p(\varepsilon_n | \sigma^2) = N(0, \sigma^2)$, with

$$y_n = f(\mathbf{X}_n, \mathbf{w}) + \varepsilon_n. \quad (1)$$

That is,

$$p(y_n | \mathbf{X}_n, \mathbf{w}, \sigma^2) = N(f(\mathbf{X}_n, \mathbf{w}), \sigma^2), \quad (2)$$

where $f(\mathbf{X}_n, \mathbf{w})$ in (1) is the prediction model for the model output, y_n , with the input \mathbf{X}_n and $\mathbf{w} = [w_1, \dots, w_N]$ is the weight vector for the RVM model.

The predicted output \hat{y} at an input \mathbf{X} in the kernel model can be represented by

$$\begin{aligned}\hat{y} &= f(\mathbf{X}, \mathbf{w}) \\ &= \sum_{i=0}^N w_i K(\mathbf{X}, \mathbf{X}_i) \\ &= \Phi \mathbf{w},\end{aligned}\quad (3)$$

where $K(\mathbf{X}, \mathbf{X}_i)$ is a basis function and Φ is the $N \times (N + 1)$ design matrix with $\Phi = [\phi(\mathbf{X}_1) \cdots \phi(\mathbf{X}_N)]^T$, wherein $\phi(\mathbf{X}_N) = [1 K(\mathbf{X}_n, \mathbf{X}_1) \cdots K(\mathbf{X}_n, \mathbf{X}_N)]^T$. In this research, radial basis function (RBF) is chosen as the basis function K because it is commonly used for modelling problems. The approach for estimating \hat{y} is to maximize the likelihood in

$$p(\mathbf{y} | \mathbf{w}, \sigma^2) = (2\pi)^{-N/2} \sigma^{-N} \exp\left\{-\frac{\|\mathbf{y} - \Phi \mathbf{w}\|^2}{2\sigma^2}\right\}. \quad (4)$$

The likelihood function in (4) is complemented by a prior over the weights $\mathbf{w} = \{w_i\}$, $i = 0$ to N , to control the complexity of the model function and avoid overfitting. The prior is a zero-mean Gaussian probability distribution and is defined over every weight w_i as follows:

$$p(\mathbf{w} | \boldsymbol{\alpha}) = (2\pi)^{-N/2} \prod_{i=0}^N \alpha_i^{1/2} \exp\left(-\frac{\alpha_i w_i^2}{2}\right). \quad (5)$$

The hyperparameters vector, $\boldsymbol{\alpha} = [\alpha_0, \dots, \alpha_N]^T$, controls how far for each weight, w_i , is allowed to deviate from zero. Consequently, using Bayes' rule, the posterior over \mathbf{w} is given as follows:

$$p(\mathbf{w} | \mathbf{y}, \boldsymbol{\alpha}, \sigma^2) = \frac{p(\mathbf{y} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha})}{p(\mathbf{y} | \boldsymbol{\alpha}, \sigma^2)}, \quad (6)$$

where $p(\mathbf{y} | \boldsymbol{\alpha}, \sigma^2)$ is the normalizing factor. $p(\mathbf{y} | \mathbf{w}, \sigma^2)$ and $p(\mathbf{w} | \boldsymbol{\alpha})$ are both Gaussian priors. The posterior mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ are as follows [17]:

$$\begin{aligned}\boldsymbol{\Sigma} &= (\mathbf{A} + \sigma^{-2} \Phi^T \Phi)^{-1}, \\ \boldsymbol{\mu} &= \sigma^{-2} \boldsymbol{\Sigma} \Phi^T \mathbf{y},\end{aligned}\quad (7)$$

where \mathbf{A} defines as $\text{diag}(\alpha_0, \dots, \alpha_N)$. In fact, the \mathbf{w} in (3) can be set to the fixed $\boldsymbol{\mu}$ for the purpose of point prediction.

Rather than extending the model to include Bayesian inference over those hyperparameters (which is analytically intractable), a most-probable point estimate, $\boldsymbol{\alpha}_{\text{MP}}$, may be found via a type II maximum likelihood procedure. That is called sparse Bayesian learning which is formulated as the local maximization with respect to $\boldsymbol{\alpha}$ of the marginal likelihood, or equivalently, its logarithm $L(\boldsymbol{\alpha})$:

$$\begin{aligned}L(\boldsymbol{\alpha}) &= \log p(\mathbf{y} | \boldsymbol{\alpha}, \sigma^2) = \log \int_{-\infty}^{\infty} p(\mathbf{y} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha}) d\mathbf{w} \\ &= -\frac{1}{2} \left[N \log 2\pi + \log |\mathbf{C}| + \mathbf{y}^T \mathbf{C}^{-1} \mathbf{y} \right],\end{aligned}\quad (8)$$

Where

$$\mathbf{C} = \sigma^2 \mathbf{I} + \Phi \mathbf{A} \Phi^T. \quad (9)$$

The covariance, $\boldsymbol{\Sigma}_{\text{MP}} = \boldsymbol{\Sigma}$, can be obtained by substituting $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{\text{MP}}$ into \mathbf{A} in (7), so that the posterior mean weight, $\boldsymbol{\mu}_{\text{MP}}$, is obtained by evaluating (7) again with $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\text{MP}}$, giving a final (posterior mean) approximator:

$$\begin{aligned}Y &= f(\mathbf{X}^*, \boldsymbol{\mu}_{\text{MP}}) \\ &= \sum_{i=0}^N \mu_{\text{MP},i} K(\mathbf{X}^*, \mathbf{X}_i) \\ &= \sum_{i=0}^N \mu_{\text{MP},i} \exp\left(-\frac{\|\mathbf{X}^* - \mathbf{X}_i\|^2}{\sigma^2}\right),\end{aligned}\quad (10)$$

where Y is the prediction of the model output with the unseen input data \mathbf{X}^* . One crucial observation is that typically the optimal values of many hyperparameters are infinite [16]. With (7), this leads to a parameter posterior infinitely peaked at zero for many weights w_i with the consequence that $\boldsymbol{\mu}_{\text{MP}}$ correspondingly comprises very few nonzero elements.

A recent analysis has showed that $L(\boldsymbol{\alpha})$ has a unique maximum with respect to α_i [16]:

$$\alpha_i = \frac{s_i^2}{q_i^2 - s_i} \quad \text{if } q_i^2 > s_i \quad (11)$$

$$\alpha_i = \infty \quad \text{if } q_i^2 \leq s_i, \quad (12)$$

and from these, it simply follows:

$$s_m = \frac{\alpha_i S_i}{\alpha_i - S_i} \quad q_i = \frac{\alpha_i Q_i}{\alpha_i - S_i}. \quad (13)$$

Note that when $\alpha_i = \infty$, $s_i = S_i$ and $q_i = Q_i$, then, it is convenient to utilize the Woodbury identity to obtain the quantities of interest:

$$S_i = \phi_i^T (\sigma^{-2} \mathbf{I}) \phi_i - \phi_i^T (\sigma^{-2} \mathbf{I}) \Phi \boldsymbol{\Sigma} \Phi^T (\sigma^{-2} \mathbf{I}) \phi_i, \quad (14)$$

$$Q_i = \phi_i^T (\sigma^{-2} \mathbf{I}) \mathbf{y} - \phi_i^T (\sigma^{-2} \mathbf{I}) \Phi \boldsymbol{\Sigma} \Phi^T (\sigma^{-2} \mathbf{I}). \quad (15)$$

The results of (14) and (15) imply that

- (1) if ϕ_i is included in the model (i.e. $\alpha_i < \infty$) yet $q_i^2 \leq s_i$, then ϕ_i can be deleted (i.e., set α_i to ∞);
- (2) if ϕ_i is excluded from the model ($\alpha_i = \infty$) and $q_i^2 > s_i$, ϕ_i can be added (i.e., set α_i to some optimal finite values).

To train and update the RVM model dynamically, a sequential learning algorithm is required. The algorithm starts with an empty model, and sequentially adds basis functions to increase the marginal likelihood, and modify their weights. Within the same principal framework, the likelihood can also be increased by deleting those basis functions which subsequently become redundant. Since this algorithm sequentially adds or deletes the basis function to or from the model, the likelihood can be continually increased by adding and deleting basis function and this mechanism make online

model update feasible. The steps of the sequential learning algorithm are shown below.

- (1) Initialize σ^2 to some sensible values (e.g., $\text{var}[\mathbf{y}] \times 0.1$) and all other α_i are notionally set to infinity.
- (2) Initialize S_n and Q_n with a single basis vector ϕ_i from (13) and (14) and compute new α_n from (11) which can be simplified as

$$\alpha_i = \frac{\|\phi_i\|^2}{\|\phi_i^T \mathbf{y}\|^2 / \|\phi_i\|^2 - \sigma^2}. \quad (16)$$

- (3) Explicitly compute Σ and μ (which are scalars initially), along with initial values of s_i and q_i for all N basis functions ϕ_i .
- (4) Select a candidate basis vector ϕ_i from the set of all N basis functions.
- (5) Compute $\theta_i = Q_i^2 - S_i$.
- (6) If $\theta_i > 0$ and $\alpha_i < \infty$ (i.e., ϕ_i is included in the model), then reestimate α_i .
- (7) If $\theta_i = 0$ and $\alpha_i = \infty$, then add ϕ_i to the model with updated α_i .
- (8) If $\theta_i < 0$ and $\alpha_i < \infty$, then delete ϕ_i from the model and set $\alpha_i = \infty$.
- (9) Estimate the noise level, update σ^2 as follows:

$$\sigma^2 = \frac{\|\mathbf{y} - \{f(\mathbf{X}_n, \mathbf{w})\}_{n=1}^N\|^2}{(N - M + \sum_i \alpha_i \Sigma_{ii})}. \quad (17)$$

- (10) Recomputed or update Σ , μ and all S_i and Q_i using (7), (13) to (15).
- (11) If converged then terminate, otherwise go to Step 4.

It has to be noticed that the RVM modelling algorithm is only a multi-input but single-output modelling method. Therefore, individual model corresponding to each output needed to be constructed. A multi-input/multioutput model is then easily be obtained by combining all the individual models.

3. Experimental Setup

Sample data sets are required for RVM training and are generally collected through experiments. In this study, the experiments were conducted on a naturally aspirated, water-cooled, 4-cylinder, direct-injection diesel engine. The specifications of the engine are shown in Table 1.

The engine was connected to an eddy-current dynamometer, and a control system was used for adjusting its speed and torque. Ultralow sulfur diesel fuel containing less than 10-ppm-wt sulfur was adopted in the test. The experimental setup is illustrated in Figure 2.

The experiments were carried out at engine speeds of 1200, 1400, 1600, 1800, and 2000 rpm and each at engine loads of 28, 70, 140, 210, and 252 Nm. For each test, the

TABLE 1: Engine specifications.

Model	Isuzu 4HF1
Type	In-line four cylinders
Maximum power	88 kW/3200 rev min ⁻¹
Maximum torque	285 Nm/1800 rev min ⁻¹
Bore \times stroke	112 mm \times 110 mm
Displacement	4334 cc
Compression ratio	19.0:1
Fuel injection timing (BTDC)	8°
Injection pump type	Bosch in-line type
Injection nozzle	Hole type (with five orifices)

volumetric flow rate of fuel was measured using a measuring cylinder and then converted into mass consumption rate, which is used to calculate the brake-specific fuel consumption (BSFC) and the brake thermal efficiency (BTE). The gaseous species in the engine exhaust including CO, CO₂, and NO_x, were measured on a continuous basis using the Anapol EU5000 exhaust gas analyzer which was suitable for measuring diesel engine emissions. The Anapol EU5000 used infrasensors for measuring CO and CO₂ concentrations and used chemical cells for measuring NO and NO₂ to obtain the NO_x concentration. The gas analyzer was calibrated with standard and zero gases before each experiment. Particulate mass concentration was measured with a tapered element oscillating microbalance (TEOM, Series 1105, Rupprecht & Patashnick Co., Inc.). The exhaust gas from the engine was diluted before passing through the TEOM with a Dekati minidiluter. The dilution ratio (DR) was evaluated based on the following equation:

$$\text{DR} = \frac{[\text{CO}_2]_{\text{exhaust}} - [\text{CO}_2]_{\text{background}}}{[\text{CO}_2]_{\text{diluted}} - [\text{CO}_2]_{\text{background}}}, \quad (18)$$

where $[\text{CO}_2]_{\text{exhaust}}$, $[\text{CO}_2]_{\text{diluted}}$, and $[\text{CO}_2]_{\text{background}}$ represent the undiluted, the diluted, and the background CO₂ concentrations, respectively. The dilution ratio was around 8 in the tests.

At each speed and load, data were recorded after the engine had reached the steady state, which was indicated by the lubricating oil temperature and the coolant temperature. For the purpose of reducing experimental uncertainties and ensuring repeatability of test data, the data were recorded continuously for 5 minutes to reduce experimental uncertainties, and each test was carried out three times. The average values were used in this research.

Based on the measured data, the following parameters are derived:

Brake thermal efficiency:

$$\text{BTE} = \eta_b = \frac{P_b}{\dot{m}_f \text{LHV}}, \quad (19)$$

Brake specific fuel consumption:

$$\text{BSFC} = \frac{\dot{m}_f}{P_b}, \quad (20)$$

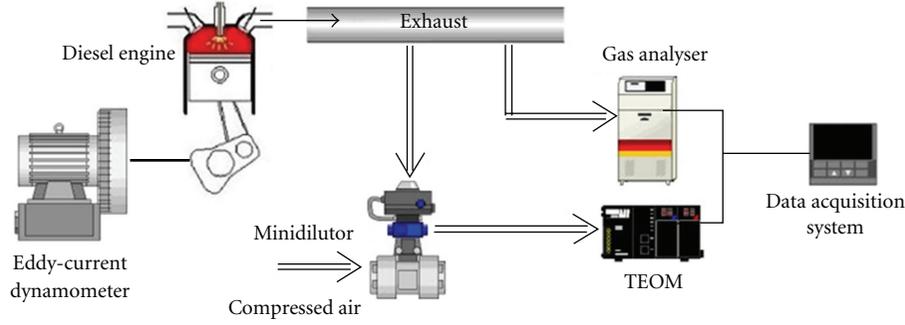


FIGURE 2: Schematic diagram of experimental setup.

where P_b is the brake power calculated from the measured torque and engine speed, \dot{m}_f is the mass flow rate of the diesel fuel and LHV is the lower heating value of the diesel fuel.

4. Application of RVM and Modelling Results

To evaluate the effectiveness of RVM, the prediction models were built based on the experimental data. As the collection of the experimental data is time consuming and costly, only 22 data sets corresponding to different load and speed settings were collected from the experiments, which are shown in Table 2. 18 sets of them were used as the training data for the model construction, and the rest 4 sets were used for model validation and testing. Actually, several weeks were required to collect the twenty-two data sets professionally. Table 3 illustrates the use of each of the data sets.

The measured parameters in each of the data sets can be basically separated into two categories, which are the input parameters and output parameters. Engine speed and engine load are the two most important independent parameters that affect the engine performance and emissions. They are included in the input parameters. The coolant controls the engine temperature so the coolant temperature is regarded as an important factor and is also treated as the input parameter. The brake-specific fuel consumption and the brake thermal efficiency represent the engine performance; thus, they are used as the output parameters. Moreover, the NO_x and particulate matter are two most serious exhaust emissions from diesel engine. Therefore, the output parameters also consist of the NO_x concentration and particle mass concentration.

The RVM modelling was implemented using MATLAB. There are three input parameters and four output parameters, indicating that four individual RVM models have to be built. Moreover, in order to have a more accurate modelling result and to prevent any input parameter from dominating the output value, the input data is conventionally normalized before training [21]. In this study, all the input values were normalized within the range $[-1, 1]$.

To verify the accuracy of the RVM model, the predicted output values is compared with the actual values from the test data sets and shown in Figures 3, 4, 5, and 6.

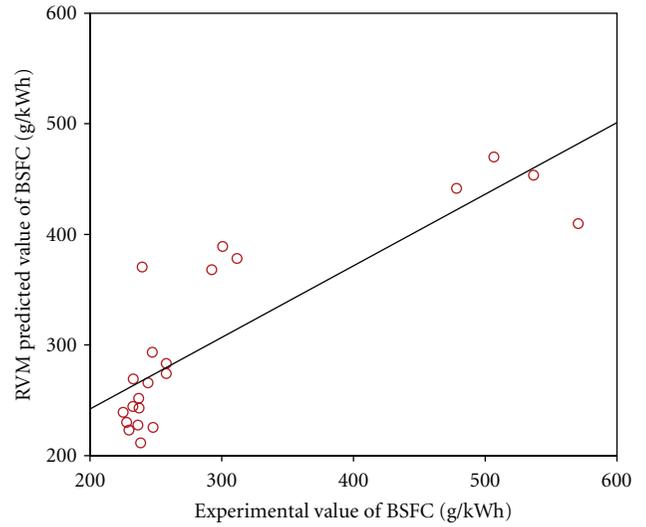


FIGURE 3: Comparison between RVM predicted values and the corresponding actual values for BSFC.

The corresponding prediction errors are presented by the mean absolute percentage error (MAPE); they were evaluated against the experimental data sets using (21). Moreover, the fraction of variance (R -squared value) is also calculated using (22) and (23). The smaller the MAPE, the better the modelling accuracy is. In addition, the best fitness of R^2 is 1

$$\text{MAPE} = \frac{1}{N_t} \sum_{k=1}^{N_t} \left| \frac{y_k - f(\mathbf{X}_k)}{y_k} \right| \times 100\%, \quad (21)$$

$$R^2 = 1 - \left(\frac{\sum_{k=1}^{N_t} (y_k - f(\mathbf{X}_k))^2}{\sum_{k=1}^{N_t} (y_k - \bar{y})^2} \right), \quad (22)$$

$$\bar{y} = \frac{1}{N_t} \sum_{k=1}^{N_t} y_k, \quad (23)$$

where \mathbf{X}_k is the k th input vectors for the prediction, $f(\mathbf{X}_k)$ is the prediction value corresponding to \mathbf{X}_k , y_k is the actual value corresponding to \mathbf{X}_k , \bar{y} is the mean of the actual value, and N_t is the number of test data points.

TABLE 2: Experimental dataset for model training and validation.

Data sets no.	Input parameters			Output parameters			
	Engine speed (rpm)	Engine load (Nm)	$T_{coolant}$ ($^{\circ}$ C)	BSFC (g/kWh)	BTE (%)	NO _x concentration (ppm)	PM mass concentration (mg/m ³)
1	1200	140	80.7	233.16	36.23	665.25	3.363
2	1200	210	82.8	225.23	37.51	900.50	9.501
3	1200	252	83.2	232.83	36.28	957.19	38.403
4	1400	28	79.3	478.57	17.65	199.84	1.233
5	1400	70	80.2	292.57	28.88	325.21	2.049
6	1400	140	80.7	236.49	35.72	610.04	3.378
7	1400	210	83.2	227.94	37.06	818.10	11.484
8	1400	252	83.7	237.09	35.63	836.57	50.906
9	1600	28	80.3	506.77	16.67	175.13	2.203
10	1600	70	80.4	300.99	28.07	303.89	3.045
11	1600	140	81.5	238.77	35.38	566.38	4.880
12	1600	210	82.3	229.92	36.74	795.66	22.049
13	1600	252	84.3	244.17	34.60	814.56	72.437
14	1800	28	80.2	537.04	15.73	170.97	1.936
15	1800	70	80.7	311.98	27.08	283.75	2.836
16	1800	140	81.4	248.01	34.07	526.49	5.323
17	1800	210	83.1	237.34	3.60	748.64	22.093
18	1800	252	85.0	258.16	32.73	730.34	115.988
19	2000	28	80.7	570.84	14.80	171.73	1.550
20	2000	70	80.8	239.71	25.62	280.93	3.418
21	2000	140	82.8	257.99	32.75	507.44	4.192
22	2000	210	83.9	247.54	34.13	683.83	26.865

TABLE 3: Data set assignment. (T and X refer to training sets and test sets, resp.)

	Engine load/torque (Nm)					
	28	70	140	210	252	
Engine speed (rpm)	1200	n/a	n/a	T	T	T
	1400	T	X	T	T	T
	1600	T	T	T	T	X
	1800	T	T	T	X	T
	2000	T	T	X	T	n/a

Table 4 summarizes the training MAPE, the MAPE over the test data sets, and the fraction of variance for each output parameter of the RVM model.

5. Comparison of RVM and ANN Modelling Results

To illustrate the advantages and superiority of the proposed RVM model, the prediction result was compared with a multilayer feed-forward neural network with backpropagation. Since multilayer feed-forward neural network is a well-known universal estimator [22] and many researches for

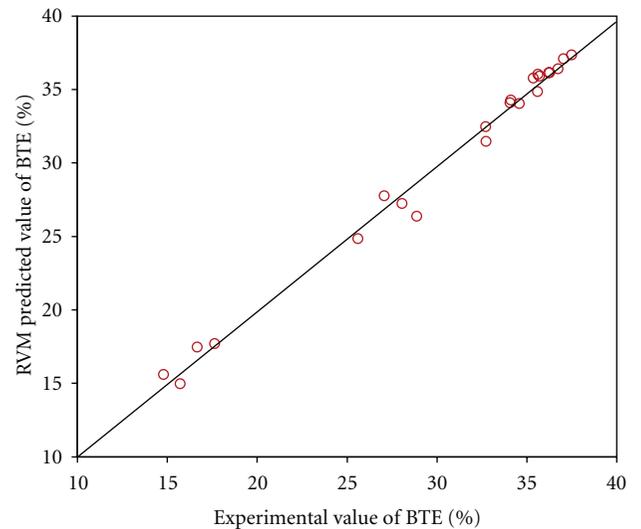
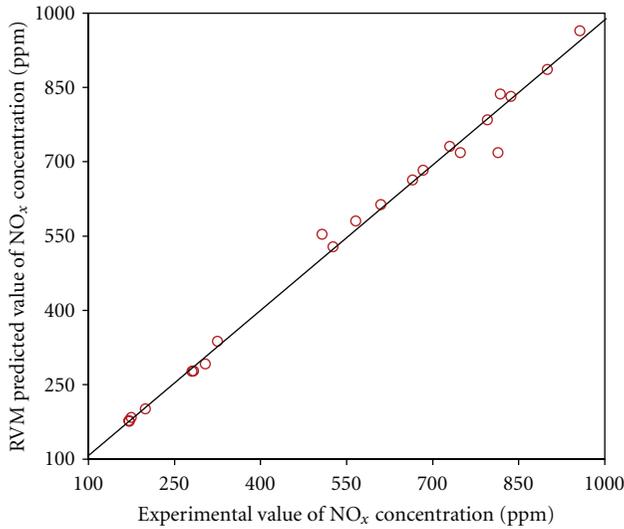


FIGURE 4: Comparison between RVM predicted values and the corresponding actual values for BTE.

diesel engine performance modelling [5–11, 13] were done based on this configuration, the results from it can be considered as a rather standard benchmark.

TABLE 4: Results of the RVM models.

Output parameters	MAPE over the training data sets (%)	MAPE over the test data sets (%)	R^2
BSFC	13.83	11.55	0.70
BTE	1.65	4.12	0.99
NO _x emission	1.61	7.17	0.99
Mass concentration	11.52	63.14	0.97
Overall average	7.15	21.50	

FIGURE 5: Comparison between RVM predicted values and the corresponding actual values for NO_x concentration.

A neural network with one hidden layer was built based on the same training data sets used for RVM modelling. The neural network consists of 3 input neurons, 20 hidden neurons, and 4 output neurons. In fact, the number of hidden nodes was determined by a trial and error analysis, varying the number of hidden neurons between 3 and 30, this burden demonstrates the ineffectiveness of the ANN approach.

The activation function used inside the hidden layer was the Tan-Sigmoid transfer function, while a pure linear filter was employed for the output layer. Levenberg-Marquardt algorithm was used as the training algorithm. The learning rate of the weight update was set to be 0.05. Figure 7 depicts the architecture of the neural network.

The same test sets were also chosen so that the RVM and ANN model can be compared reasonably. The prediction accuracy of each output in the ANN model is illustrated in Figures 8, 9, 10, and 11 and Table 5.

Tables 4 and 5 show that the RVM outperforms the ANN by about 36.45% in terms of average MAPE under the same test sets. The relatively high training MAPE of the ANN shows that the data sets is not sufficient for building such a highly nonlinear model. Furthermore, only one initial

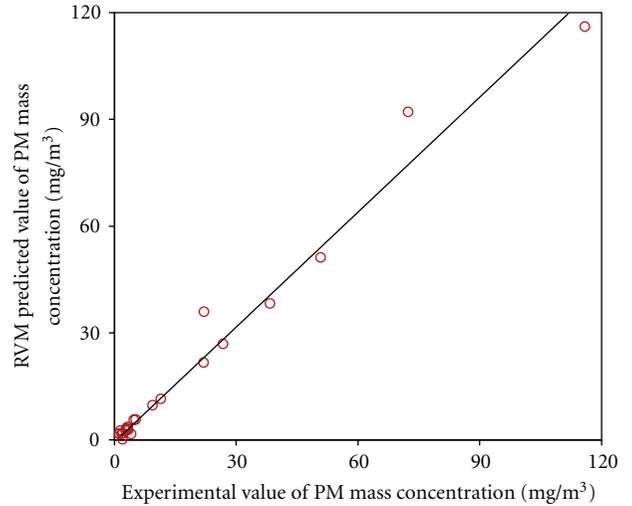


FIGURE 6: Comparison between RVM predicted values and the corresponding actual values for PM mass concentration.

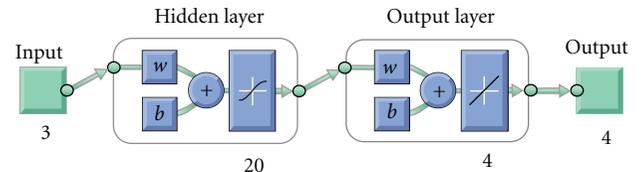


FIGURE 7: Architecture of the neural network.

value σ of the basis width is required by RVM, while the learning rate, number of hidden layers, and number of hidden neurons are required in ANN, which means a grid of guessed values for these parameters have to be prepared.

The MAPEs of both RVM and ANN for predicting the mass concentration of particulates are relatively large as compared to the other output parameter. This is because the variation of the mass concentration ranges from 0 to $12 \times 10^4 \mu\text{g}/\text{m}^3$. Actually, the RVM model tries to fit a function for the whole range rather than focusing on the low end of range, which is seen by the R-squared value of 0.97. In contrast, the ANN model tends to concentrate at the low end of the value. As a result, the R-squared value for the ANN model is only 0.02, which is unacceptable.

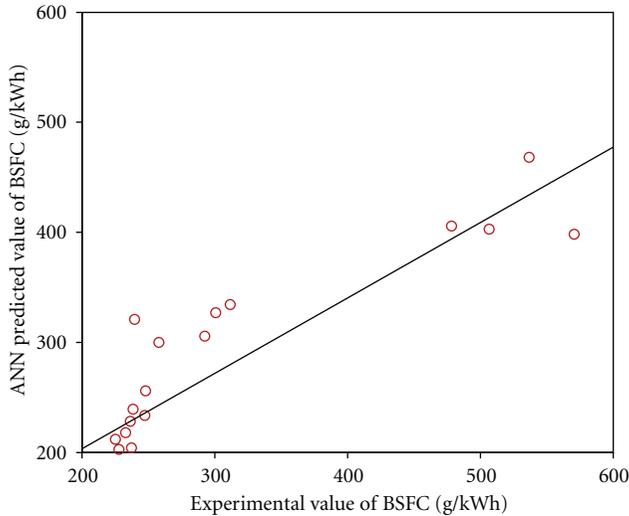


FIGURE 8: Comparison between ANN predicted values and the corresponding actual values for BSFC.

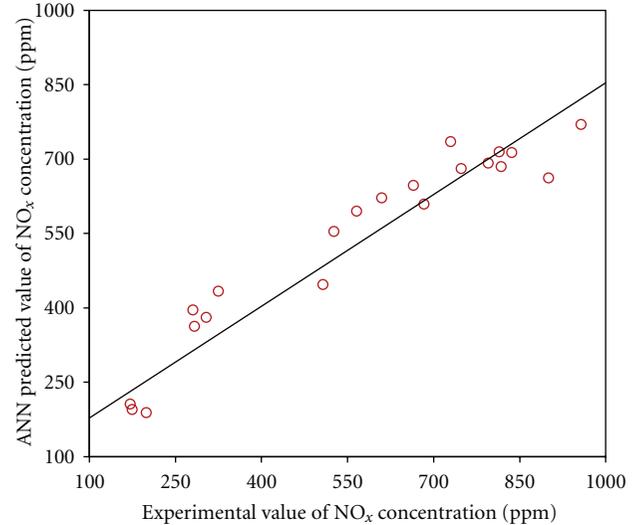


FIGURE 10: Comparison between ANN predicted values and the corresponding actual values for NO_x concentration.

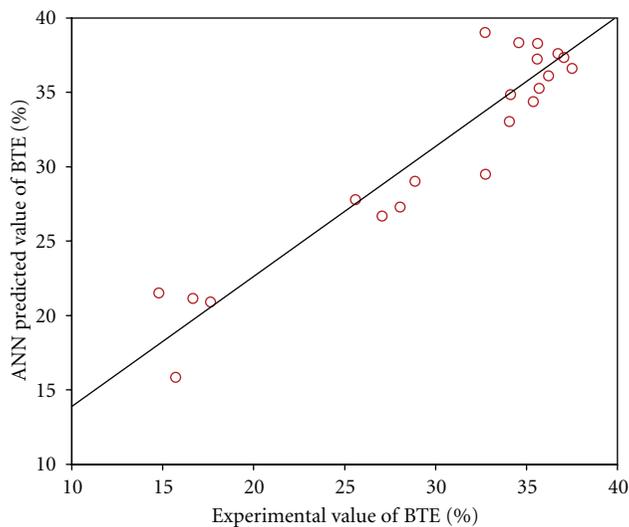


FIGURE 9: Comparison between ANN predicted values and the corresponding actual values for BTE.

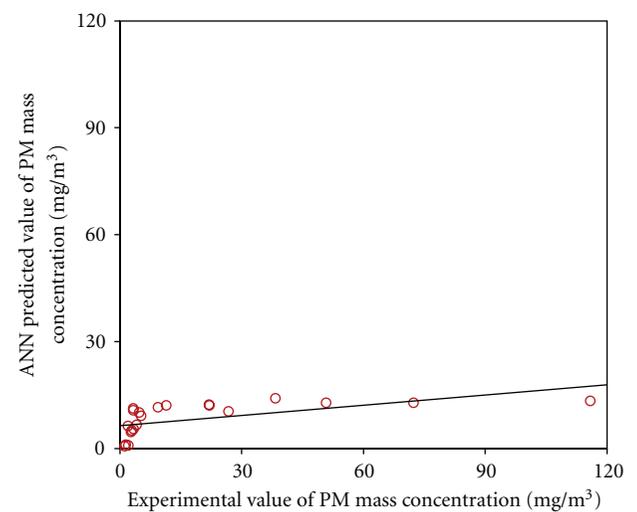


FIGURE 11: Comparison between ANN predicted values and the corresponding actual values for PM mass concentration.

Overall, the prediction accuracy of RVM with a small amount of training data is satisfactory.

6. Conclusions

This research is the first attempt at applying RVM to model the diesel engine performance and emission characteristics of NO_x and particulate matter under the condition of rare data. Although the combustion process of the diesel engine is unknown, the RVM model has successfully demonstrated the relation between the controllable factors, which are the engine speeds, engine loads, and coolant temperature, and the output variables, including the brake-specific fuel consumption, brake thermal efficiency, NO_x emission, and particulate mass concentrations. Experimental results show

that the RVM model is still acceptable even if the data sets are few. It is believed that more data sets can improve the accuracy of the model.

Furthermore, the RVM model is also compared with an ANN model. The results indicate that the average accuracy of the RVM model is higher than that of the ANN model by about 36.45%, implying that RVM is superior to ANN.

With the proposed RVM model, experimental efforts can be reduced significantly as the performance and emissions of the diesel engine can be predicted easily. By applying this RVM model as a virtual sensor on diesel vehicles, the exhaust emissions can be controlled more effectively by incorporating with some advanced control algorithms, such as model predictive control. The study of model predictive diesel emission control based on RVM model will

TABLE 5: Results of the ANN model.

Output parameters	MAPE over the training data sets (%)	MAPE over the test data sets (%)	R^2
BSFC	14.13	14.12	0.70
BTE	8.66	6.42	0.86
NO _x emission	18.15	16.64	0.85
Mass concentration	87.42	98.12	0.02
Overall average	32.09	33.83	

be considered as a future work. Since RVM can also perform online model update, the applications of RVM to online system modelling and online control will also be explored in the future.

Acknowledgments

The research is supported by the University of Macau Research Grant, Grant no. MYRG149(Y2-L2)-FST11-WPK and the short-term visiting scholar programme of University of Macau. The authors would also like to thank the support from the Hong Kong Polytechnic University and the technician, Mr. Wong, Hang Cheong, of the Automotive Engineering Laboratory of University of Macau.

References

- [1] D. W. Dockery and C. A. Pope, "Acute respiratory effects of particulate air pollution," *Annual Review of Public Health*, vol. 15, pp. 107–132, 1994.
- [2] J. E. Penner, C. C. Chuang, and K. Grant, "Climate forcing by carbonaceous and sulfate aerosols," *Climate Dynamics*, vol. 14, no. 12, pp. 839–851, 1998.
- [3] M. Z. Jacobson, "Strong radiative heating due to the mixing state of black carbon in atmospheric aerosols," *Nature*, vol. 409, no. 6821, pp. 695–697, 2001.
- [4] W. Knecht, "Diesel engine development in view of reduced emission standards," *Energy*, vol. 33, no. 2, pp. 264–271, 2008.
- [5] M. S. Rawlins, *Diesel engine performance modelling using neural networks*, Ph.D. thesis, Department of Mechanical Engineering, Durban Institute of Technology, Durban, South Africa, 2005.
- [6] V. Çelik and E. Arcaklıoğlu, "Performance maps of a diesel engine," *Applied Energy*, vol. 81, no. 3, pp. 247–259, 2005.
- [7] M. Canakci, A. Erdil, and E. Arcaklıoğlu, "Performance and exhaust emissions of a biodiesel engine," *Applied Energy*, vol. 83, no. 6, pp. 594–605, 2006.
- [8] Shivakumar, P. Srinivasa Pai, and B. R. Shrinivasa Rao, "Artificial Neural Network based prediction of performance and emission characteristics of a variable compression ratio CI engine using WCO as a biodiesel at different injection timings," *Applied Energy*, vol. 88, no. 7, pp. 2344–2354, 2011.
- [9] H. Oğuz, I. Saritas, and H. E. Baydan, "Prediction of diesel engine performance using biofuels with artificial neural network," *Expert Systems with Applications*, 2010.
- [10] C. Brace, "Prediction of diesel engine exhaust emission using artificial neural networks," in *Proceedings of the Neural Networks in Systems Design' IMechE Seminar S591*, Solihull, UK, 1998.
- [11] M. Traver, R. Atkinson, and C. Atkinson, "Neural network-based diesel engine emissions prediction using in-cylinder combustion pressure," SAE Paper 1999-01-1532, 1999.
- [12] Z. T. Liu and S. M. Fei, "Study of CNG/diesel dual fuel engine's emissions by means of RBF neural network," *Journal of Zhejiang University*, vol. 5, no. 8, pp. 960–965, 2004.
- [13] M. Ghazikhani and I. Mirzaii, "Soot emission prediction of a waste-gated turbo-charged DI diesel engine using artificial neural network," *Neural Computing and Applications*, vol. 20, no. 2, pp. 303–308, 2011.
- [14] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice-Hall, Upper Saddle River, NJ, USA, 1999.
- [15] P. K. Wong, L. M. Tam, and K. Li, "Automotive engine power performance tuning under numerical and nominal data," *Control Engineering Practice*, vol. 20, no. 3, pp. 300–314, 2011.
- [16] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research*, vol. 1, no. 3, pp. 211–244, 2001.
- [17] M. E. Tipping and A. C. Faul, "Fast marginal likelihood maximisation for sparse Bayesian method," in *Proceedings of 9th International Workshop on Artificial Intelligence and Statistics*, Key West, Fla, USA, January 2003.
- [18] J. Yuan, K. Wang, T. Yu, and M. Fang, "Integrating relevance vector machines and genetic algorithms for optimization of seed-separating process," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 7, pp. 970–979, 2007.
- [19] W. Caesarendra, A. Widodo, and B. S. Yang, "Application of relevance vector machine and logistic regression for machine degradation assessment," *Mechanical Systems and Signal Processing*, vol. 24, no. 4, pp. 1161–1171, 2010.
- [20] M. Han and Y. Zhao, "Dynamic control model of BOF steel-making process based on ANFIS and robust relevance vector machine," *Expert Systems with Applications*, vol. 38, no. 12, pp. 14786–14798, 2011.
- [21] D. Pyle, *Data Preparation for Data Mining*, Morgan Kaufmann Publishers, 1999.
- [22] C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, New York, NY, USA, 1995.