

ตัวแบบผสมสำหรับการพยากรณ์คุณภาพอากาศ กรณีศึกษา ข้อมูลคุณภาพอากาศในประเทศไทย

Hybrid model for air quality data prediction Case study: air quality data in Thailand

ชยกร สมศิลา (Chayakorn Somsila)* สิริภัทร เชื้อชาญวัฒนา (Sirapat Chiewchanwattana)**

คำรณ สุนัตติ (Khamron Sunat)***

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ศึกษาหาตัวแบบผสมในการพยากรณ์คุณภาพอากาศระหว่างวิธีของบ็อกซ์และเจนกินส์กับโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ โดยใช้ข้อมูลคุณภาพอากาศเฉลี่ยรายเดือนก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) ก๊าซไนโตรเจนไดออกไซด์ (NO_2) ก๊าซคาร์บอนมอนอกไซด์ (CO) และก๊าซโอโซน (O_3) ของสถานีตรวจวัดคุณภาพอากาศอนามัยมาตาพุด อำเภอมาตาพุด จังหวัดระยอง และเปรียบเทียบประสิทธิภาพการพยากรณ์ระยะสั้นและระยะยาว ตัวแบบบ็อกซ์และเจนกินส์ ตัวแบบโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ และตัวแบบผสมระหว่างวิธีบ็อกซ์และเจนกินส์กับโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ พบว่า ตัวแบบผสมระหว่างวิธีของบ็อกซ์และเจนกินส์ร่วมกับโครงข่ายประสาทเทียมให้ผลการพยากรณ์ที่ถูกต้องที่สุดทั้งในการพยากรณ์ระยะสั้นและระยะยาว

ABSTRACT

This research study to find hybrid model between Box & Jenkins and feed forward backpropagation artificial neural network by air-quality data average monthly of Sulfur dioxide (SO_2), Nitrogen dioxide (NO_2), Carbon monoxide (CO) and Ozone (O_3) from Map Ta Phut stations, Amphor Map Ta Phut, Rayong Province. And compare performance predict short term and long term, Box & Jenkins model, feed forward backpropagation artificial neural network model and hybrid model, found hybrid model is the most accurate predict in short-term and long term.

คำสำคัญ : ตัวแบบผสม บ็อกซ์และเจนกินส์ โครงข่ายประสาทเทียม

Key Words : Hybrid, Box & Jenkins, Neural network

* นักศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

** ผู้ช่วยศาสตราจารย์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

*** อาจารย์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีมหานคร

บทนำ

ปัญหาหมอกพิษทางอากาศเป็นปัญหาสำคัญปัญหาหนึ่งที่พบในประเทศไทย โดยเฉพาะในเขตเมืองใหญ่ ๆ ที่มีการเจริญเติบโตทางด้านอุตสาหกรรม เช่น กรุงเทพมหานคร ระยอง ชลบุรี เป็นต้น โดยสถานการณ์มลพิษทางอากาศในปี 2551 ที่ผ่านมา ได้สรุปไว้ว่า สถานการณ์คุณภาพอากาศในภาพรวมแม้จะดีขึ้นกว่าเดิม แต่ก็ยังตรวจพบมลพิษในอากาศที่เกินมาตรฐาน เช่น การพบปริมาณโอโซนที่เกินมาตรฐานในกรุงเทพมหานครและเขตปริมณฑล หรือการตรวจพบก๊าซซัลเฟอร์ไดออกไซด์ ก๊าซไนโตรเจนออกไซด์ และก๊าซคาร์บอนมอนอกไซด์ที่เกินมาตรฐาน และพบว่าเกิดข้อร้องเรียนเกี่ยวกับปัญหาหมอกพิษทางอากาศและเสียง สูงที่สุดเป็นอันดับหนึ่ง คือ ร้อยละ 82 ของปัญหาหมอกพิษที่พบทั้งหมด (มติชนออนไลน์, 2551) จึงมีการประกาศค่าปริมาณสารมลพิษที่ยอมให้ปล่อยสู่บรรยากาศเพื่อควบคุมสารมลพิษหลักจำนวน 7 ชนิด คือ ก๊าซคาร์บอนมอนอกไซด์ ก๊าซไนโตรเจนไดออกไซด์ ก๊าซซัลเฟอร์ไดออกไซด์ ฝุ่นรวม ตะกั่ว ฝุ่นขนาดเล็กและโอโซน (นันทวัฒน์, 2551) จากที่กล่าวมานี้ถ้าเราสามารถที่จะประเมินเกี่ยวกับคุณภาพอากาศหรือสามารถประมาณการณได้ว่าสภาพอากาศที่จะเกิดขึ้นในอนาคตเป็นอย่างไร ย่อมมีผลดีกับการดำรงชีวิตของมนุษย์ เพราะจะทำให้ทราบถึงสภาวะมลภาวะหรือปัญหาที่จะเกิดขึ้น ดังนั้นความต้องการทราบข้อมูลในอนาคตหรือการคาดการณ์ข้อมูลล่วงหน้าหรือการพยากรณ์ สามารถที่จะช่วยคาดคะเนหรือช่วยในการประเมิน ทำให้รู้แนวโน้มของเหตุการณ์ในอนาคต ส่งผลถึงการตัดสินใจหรือการวางแผนหรือหาทางป้องกันสำหรับเหตุการณ์ในอนาคตได้

ในการศึกษาค้นคว้าครั้งนี้ผู้วิจัยมุ่งเน้นที่จะทำการพยากรณ์ข้อมูลคุณภาพอากาศจากสถานีตรวจวัด โดยใช้ การพยากรณ์อนุกรมเวลา เป็นวิธีการพยากรณ์เชิงปริมาณรูปแบบหนึ่ง ด้วยการนำข้อมูลในอดีตที่ทำการเก็บรวบรวมมาเป็นระยะเวลาหนึ่งมาทำการศึกษา

ลักษณะการเปลี่ยนแปลงของข้อมูลว่าข้อมูลมีการเปลี่ยนแปลงอย่างไร มีการเคลื่อนไหวเป็นแบบใด มาสร้างตัวแบบทางคณิตศาสตร์ขึ้นเพื่อใช้เป็นตัวแบบในการพยากรณ์ เนื่องจากข้อมูลในอดีตจะส่งผลต่อการเคลื่อนไหวของข้อมูลในอนาคต (ทรงศิริ, 2549) มีนักวิจัยจำนวนมาก ได้ทำการศึกษาเกี่ยวกับการหาตัวแบบทางคณิตศาสตร์ที่ใช้ในการพยากรณ์ เช่น Luis and Robles (2008) นำตัวแบบบอซและเจนกินส์ (Autoregressive integrated moving average: ARIMA) และโครงข่ายประสาทเทียม (Artificial neural network: ANN) มาผสมกัน (Hybrid) เพื่อใช้ในการพยากรณ์สภาพอากาศในเมือง Temuco ประเทศชิลี (เป็นสถานที่ที่มีไฟป่าเกิดขึ้นก่อให้เกิดมลพิษทางอากาศในช่วงหน้าหนาว) Hashemloo (2008) ได้ศึกษาเปรียบเทียบการพยากรณ์อุณหภูมิของพื้นผิวหน้าถนนในโตรอนโต ประเทศแคนาดา โดยใช้การพยากรณ์อนุกรมเวลาและแบบจำลองโครงข่ายประสาทเทียม Nunnari และ Cannav' (2005) การหาตัวแบบโครงข่ายประสาทเทียม แบบแบพร้อนกลับ (Backpropagation) ที่เหมาะสมสำหรับการพยากรณ์ข้อมูลอนุกรมเวลาดำวยในการศึกษาข้อมูลคุณภาพอากาศในอิตาลี Chelani และ Devotta (2006) ใช้ตัวแบบผสมในการพยากรณ์ก๊าซไนโตรเจนไดออกไซด์ ในเมือง Delhi จากงานวิจัยที่กล่าวมานี้ จะเห็นว่าได้มีการพัฒนาวิธีการผสม (Hybrid) ระหว่างวิธีการพยากรณ์ข้อมูลส่วนที่เหมาะสมกับฟังก์ชันเชิงเส้น (linear function) รวมกับวิธีการพยากรณ์ข้อมูลส่วนที่เหมาะสมกับฟังก์ชันไม่เชิงเส้น (non-linear function) ของข้อมูลอนุกรมเวลา (Zhang, 2003) เพื่อให้การพยากรณ์มีความถูกต้องแม่นยำมากยิ่งขึ้น และพบว่าใช้โครงข่ายประสาทเทียมพยากรณ์ข้อมูลส่วนที่เป็นฟังก์ชันไม่เชิงเส้นของข้อมูลอนุกรมเวลาและวิธีของบอซและเจนกินส์ วิธีปรับเรียบเอ็กซ์โปเนนเชียล วิธีถดถอยเชิงพหุ ใช้พยากรณ์ข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้นของข้อมูลอนุกรมเวลา

ดังนั้นในการศึกษาค้นคว้าครั้งนี้จะทำการหาตัวแบบผสมสำหรับการพยากรณ์ข้อมูลคุณภาพอากาศเฉลี่ยรายเดือนของประเทศไทย โดยใช้ตัวแบบผสมระหว่างวิธีของ

บอซและเจนกินส์ซึ่ง “เป็นวิธีการพยากรณ์ระยะสั้นที่ดี” (ทรงศิริ, 2549) พยากรณ์ข้อมูลอนุกรมเวลาที่เป็นฟังก์ชันเชิงเส้นรวมกับวิธีโครงข่ายประสาทเทียม แบบ feed forward backpropagation พยากรณ์ข้อมูลอนุกรมเวลาส่วนที่เป็นฟังก์ชันไม่เชิงเส้น และใช้โปรแกรม SPSS ในการหาตัวแบบ ARIMA และโปรแกรม Matlab หาตัวแบบ ANN

วัตถุประสงค์การวิจัย

1. หาตัวแบบการพยากรณ์แบบผสมระหว่างวิธีของบอซและเจนกินส์ และวิธีโครงข่ายประสาทเทียม แบบ feed forward backpropagation ของข้อมูลคุณภาพอากาศเฉลี่ยรายเดือนก๊าซซัลเฟอร์ไดออกไซด์ ก๊าซไนโตรเจนไดออกไซด์ ก๊าซคาร์บอนมอนอกไซด์ และโอโซนของสถานีตรวจวัดคุณภาพอากาศอนามัยมาตาศุด ตำบลมาตาศุด อำเภอเมือง จังหวัดระยอง
2. ทดสอบและวัดประสิทธิภาพการพยากรณ์ข้อมูลคุณภาพอากาศเฉลี่ยรายเดือนด้วยตัวแบบผสมระหว่างวิธีของบอซและเจนกินส์และวิธีโครงข่ายประสาทเทียม แบบ feed forward back propagation

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

อนุกรมเวลา

ข้อมูลที่เก็บรวบรวมไว้ในระยะเวลาหนึ่งที่มีช่วงห่างเท่ากัน และมีองค์ประกอบการผันแปร 4 ส่วน ได้แก่ แนวโน้ม ฤดูกาล การเปลี่ยนแปลงตามวัฏจักร และการเปลี่ยนแปลงที่เกิดจากเหตุการณ์ปกติ (ทรงศิริ, 2549)

การประมาณค่าที่สูญหาย

ในงานวิจัยนี้ใช้เบสส์กับสไลด์ piecewise cubic hermite Interpolating polynomial (pchip) (The mathworks, 2007) โปรแกรม Matlab ซึ่งเป็นการประมาณค่าระหว่างจุดสองจุดด้วยพหุนามกำลังสามช่วยในการประมาณค่าข้อมูลที่สูญหาย

บอซและเจนกินส์(Box and Jenkins) (ตัวแบบ ARIMA)

เป็นวิธีการพยากรณ์ที่รวมเอาตัวแบบการถดถอยในตนเองซึ่งเป็นค่าสังเกตในอดีต (Autoregressive หรือ AR) และตัวแบบค่าเฉลี่ยเคลื่อนที่ซึ่งเป็นความคลาดเคลื่อนสุ่มในอดีต (Moving average หรือ MA) เข้าด้วยกัน โดยเป็นการหาสมการเพื่อแสดงความสัมพันธ์ระหว่างชุดข้อมูลกับเวลาที่เปลี่ยนแปลงไป โดยตัวแบบ ARMA เมื่อรวมกับกระบวนการ Integrated รวมกันแล้วได้ตัวแบบดังสมการต่อไปนี้

$$\Delta_d y_t = \delta + \phi_1 \Delta_d y_{t-1} + \dots + \phi_p \Delta_d y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (1)$$

เมื่อ y_t คือ ค่าสังเกตของอนุกรมเวลาที่เวลา t

d คือ จำนวนครั้งการหาผลต่างที่ทำให้อนุกรมเวลาคงที่

p คือ อันดับของ autoregressive

q คือ อันดับของ moving average

δ คือ ค่าคงที่

t คือ เวลา

Δ_d คือ ผลต่างอันดับที่ d

$\phi_1, \dots, \phi_p$ คือ พารามิเตอร์ของ autoregressive

$\theta_1, \dots, \theta_q$ คือ พารามิเตอร์ของ moving average

ε_t คือ ค่าความคลาดเคลื่อน ณ เวลา t ภายใต้ข้อสมมติที่ว่าความคลาดเคลื่อนที่แต่ละเวลาเป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน โดยมี การแจกแจงปกติที่มีค่าเฉลี่ยเป็นศูนย์ และความแปรปรวนคงที่ ขั้นตอนการสร้างตัวแบบมี 4 ขั้นตอน ดังนี้

1. กำหนดตัวแบบ อนุกรมเวลาที่นำมาสร้างตัวแบบต้องสแตชันนารี (stationary) คือความแปรปรวนและค่าเฉลี่ยคงที่ ถ้าอนุกรมเวลาไม่สแตชันนารี ต้องทำการแปลงอนุกรมเวลานั้น เช่น ค่าเฉลี่ยไม่คงที่ แปลงด้วยการหาผลต่าง ความแปรปรวนไม่คงที่ แปลงอนุกรมเวลาด้วยฟังก์ชันลอการิทึม เป็นต้น แล้วนำอนุกรมเวลาไปพล็อตกราฟคอเรลโรแกรม (correlogram) เพื่อพิจารณาดูเปรียบเทียบกับแผนภาพคอเรลโรแกรมตัวแบบ (ทรงศิริ, 2549) จะได้รูปแบบ ARMA(p,q)

2. การประมาณค่าของตัวแบบ โดยให้ค่าประมาณที่ได้มีความคลาดเคลื่อนน้อยที่สุด

3. การตรวจสอบตัวแบบ พิจารณาค่าที่เหลือ (residual) ที่ได้จากตัวแบบสอดคล้องกับสมมติฐานหรือไม่ ถ้าไม่สอดคล้องต้องกลับไปเริ่มหาตัวแบบใหม่

4. นำตัวแบบที่ผ่านการตรวจสอบแล้ว ไปใช้ในการพยากรณ์

การแปลงฟูเรียร์ Fast Fourier Transform (FFT) (The Mathworks, 2007)

เป็นวิธีที่ใช้ในการเปลี่ยนรูปคลื่นสัญญาณทางเวลาไปเป็นส่วนประกอบกันทางความถี่หรือสเปกตรัมของสัญญาณนั้นๆ ในงานวิจัยนี้ได้ใช้ประโยชน์จากแบล็กบ็อกซ์ในการหาคาบของข้อมูลอนุกรมเวลาเพื่อกำหนด input node ในโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (ตัวแบบ ANN)

เป็นตัวแบบทางคณิตศาสตร์ที่ถูกสร้างขึ้นเพื่อเลียนแบบการทำงานของสมองมนุษย์ โครงข่ายประสาทเทียมที่เลือกใช้นี้เป็นโครงข่ายประสาทเทียมแบบต้องมีผู้สอน (supervised learning)

สถาปัตยกรรมและการทำงานของโครงข่ายประสาทเทียม

ประกอบด้วยสามชั้นหลัก ได้แก่ ชั้นนำข้อมูลเข้า (input layer) ทำหน้าที่รับข้อมูลและส่งข้อมูลไปสู่ชั้นถัดไป ชั้นซ่อน (hidden layer) ทำหน้าที่รับข้อมูลจากชั้นก่อนหน้านี้นี้มาคูณกับค่าน้ำหนัก (weight) ดังสมการ และประมวลผลออกด้วยฟังก์ชันกระตุ้น (activation function) ที่อยู่ในชั้นนี้ และได้ข้อมูลออกส่งไปที่ชั้นนำข้อมูลออก (output layer) ซึ่งก็คือค่าพยากรณ์ มาคำนวณเปรียบเทียบกับข้อมูลค่าจริง ทำการคำนวณค่าคลาดเคลื่อน (error) ที่เกิดขึ้น แล้วส่งค่าคลาดเคลื่อนของชั้นผลลัพธ์ย้อนกลับไปชั้นต่างๆ ของโครงข่าย เพื่อปรับน้ำหนักใหม่เพื่อลดความคลาดเคลื่อน และทำการหยุดเมื่อค่าคลาดเคลื่อนลดลงน้อยกว่าที่กำหนด หรือจำนวนรอบของการฝึกดำเนิน

ไปจนถึงค่าที่ตั้งไว้ การเกิด over fitting (พนิดาและคณะ, 2009) เป็นปรากฏการณ์ที่ค่าคลาดเคลื่อนของการฝึกสอนของโครงข่ายมีค่าน้อยมาก แต่เมื่อใช้โครงข่ายประสาทเทียมเดิมกับข้อมูลใหม่กลับมีค่าคลาดเคลื่อนที่สูง แสดงให้เห็นว่าโครงข่ายประสาทเทียมสามารถจำโครงข่ายมาฝึกสอนแต่ไม่สามารถเรียนรู้สถานการณ์ใหม่ได้ ในงานวิจัยนี้ได้ใช้เทคนิค early stopping ในการแก้ปัญหา

ตัวแบบผสม

Zhang (2003) ได้นำเสนอว่า ตัวแบบนี้มีข้อสมมติว่าอนุกรมเวลามีส่วนประกอบแบบเชิงเส้น L_t และส่วนประกอบแบบไม่เชิงเส้น N_t นั่นคือตัวแบบจะมีลักษณะเป็นดังสมการ

$$Y_t = L_t + N_t + \varepsilon_t \quad (4)$$

เมื่อ Y_t คือ ค่าสังเกตของอนุกรมเวลาที่เวลา t

L_t คือ ส่วนประกอบแบบเชิงเส้น

N_t คือ ส่วนประกอบแบบไม่เชิงเส้น

ε_t คือ ความคลาดเคลื่อนสุ่มที่เวลา t

ขั้นตอนวิธีของตัวแบบผสมระหว่าง ARIMA และ ANN มีดังนี้

(1) ใช้ตัวแบบ ARIMA วิเคราะห์ข้อมูลอนุกรมเวลา เพื่อประมาณส่วนประกอบที่เป็นเชิงเส้น $\hat{L}_t$

(2) หาส่วนเหลือจากตัวแบบ ARIMA นั่นคือ $\varepsilon_t = Y_t - \hat{L}_t$ (ส่วนที่เหลือ ε_t เป็นค่าสังเกตที่ถูกจำกัดส่วนฟังก์ชันเชิงเส้นออกแล้ว)

(3) ใช้วิธี ANN กับส่วนเหลือ (ε_t) ที่ได้ในขั้นที่ 2 เพื่อประมาณส่วนประกอบที่ไม่ใช่เชิงเส้น $\hat{N}_t$

(4) รวมค่าประมาณส่วนประกอบทั้ง 2 ส่วน เพื่อใช้ประมาณค่าอนุกรมเวลา $\hat{Y}_t = \hat{L}_t + \hat{N}_t$

ข้อมูลที่ใช้ทดลอง

ข้อมูลคุณภาพอากาศเฉลี่ยรายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2540 ถึงเดือนธันวาคม พ.ศ. 2551 ของสถานีตรวจวัดคุณภาพอากาศอนามัยมาบตาพุด ตำบลมาบตาพุด อำเภอเมือง จังหวัดระยอง (รหัสสถานี 29T) ประกอบด้วยข้อมูลคุณภาพอากาศเฉลี่ยรายเดือนของก๊าซซัลเฟอร์ได

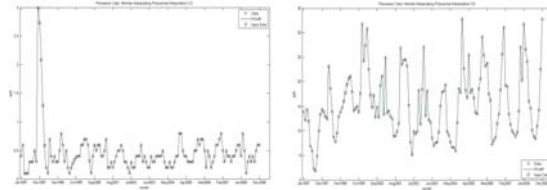
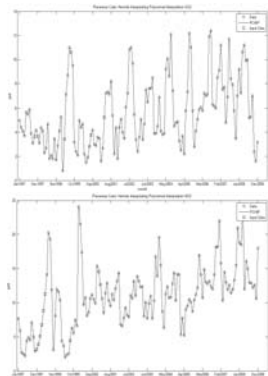
ออกไซด์ (SO_2) ก๊าซไนโตรเจนไดออกไซด์ (NO_2) ก๊าซคาร์บอนมอนนอกไซด์ (CO) และก๊าซโอโซน (O_3) ซึ่งในแต่ละก๊าซจะมีข้อมูลเท่ากับ 144 ข้อมูล

วิธีการทดลองและผลการทดลอง

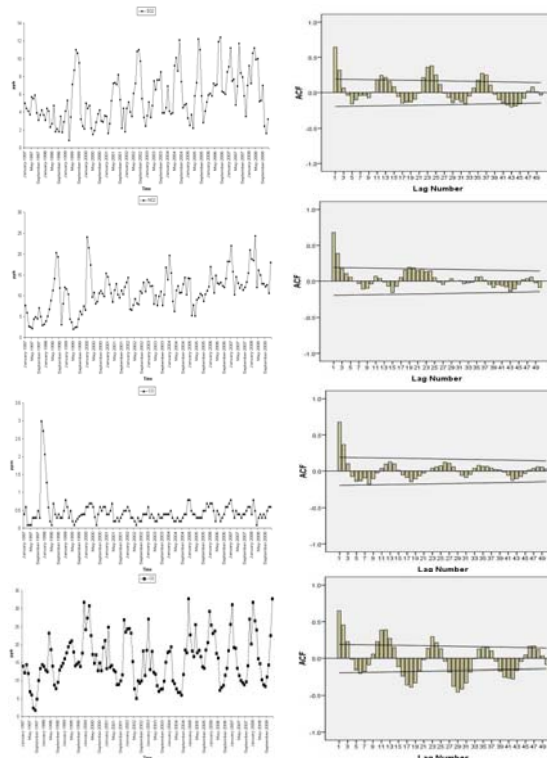
1. เตรียมข้อมูล พบว่าข้อมูลไม่สมบูรณ์จึงประมาณค่าส่วนที่สูญหายของข้อมูลให้ครบ ด้วยแบล็กบ็อก piecewise cubic hermite interpolating polynomial (pchip) ใน matlab ดังภาพที่ 1

2. นำข้อมูลมาพิจารณาคุณลักษณะส่วนประกอบของอนุกรมเวลา เช่น แนวโน้ม ฤดูกาล สามารถดูกราฟฟังก์ชันสหสัมพันธ์ (Autocorrelation function: ACF) ประกอบด้วย และเพื่อความชัดเจนควรหาผลต่างลำดับที่ 1 ประกอบ ดังภาพที่ 2 พบว่าข้อมูล SO_2 และ O_3 มีการแปรผันตามฤดูกาล กราฟ NO_2 มีแนวโน้ม และ CO มีการเปลี่ยนแปลงที่เกิดจากเหตุการณ์ปกติ สรุปได้ว่าข้อมูลทุกก๊าซเป็นข้อมูลอนุกรมเวลา

3. ตัวแบบ ARIMA แบ่งข้อมูลออกเป็น 2 ชุด ได้แก่ ชุดใช้หาตัวแบบจำนวน 108 ข้อมูล และชุดใช้ทดสอบจำนวน 36 ข้อมูล



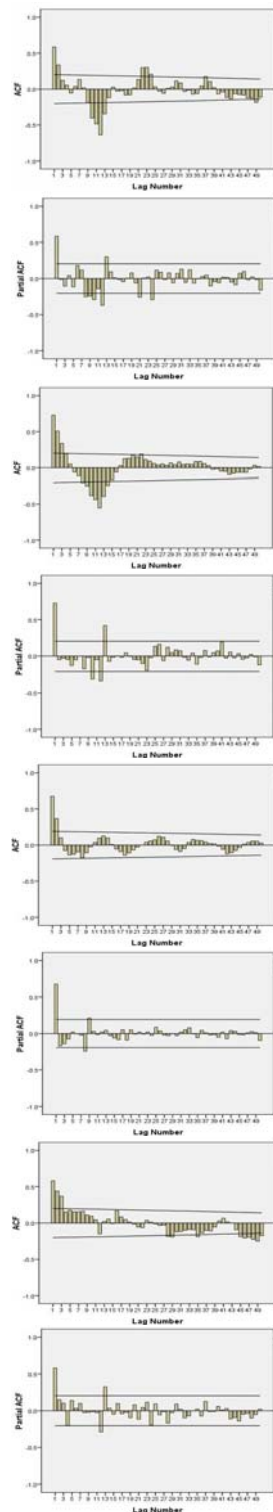
ภาพที่ 1 ใช้แบล็กบ็อก pchip ประมาณค่าที่สูญหายของก๊าซ SO_2 , NO_2 , CO และ O_3 เรียงลำดับ จากซ้ายไปขวาและบนลงล่าง



ภาพที่ 2 กราฟข้อมูลคุณภาพอากาศเฉลี่ยรายเดือนกับเวลา และกราฟ ACF ของ SO_2 , NO_2 , CO และ O_3 เรียงจากบนลงล่าง

1) เมื่อทำอนุกรมเวลาสเตชันนารีแล้ว ก็พิจารณากราฟฟังก์ชันตัวอย่างกระบวนการสเตชันนารี (ทรงศิริ, 2549) Sample autocorrelation function (SACF) และ Sample partial autocorrelation function (SPACF) ดังภาพที่ 3 เทียบกับตัวอย่างรูปแบบของ SACF และ SPACF ที่พบบ่อย (ทรงศิริ, 2549) แล้วกำหนดตัวแบบได้ดังนี้ SO_2 ตัวแบบคือ $\text{ARIMA}(1,0,0)(0,1,1)_n$ NO_2 ตัวแบบคือ $\text{ARIMA}(1,0,0)(1,1,1)_n$ CO ตัวแบบคือ

ARIMA(1,0,0)(0,0,0)₁₂ และ O₃ ตัวแบบของคือ
ARIMA(0,0,3)(1,1,0)₁₂

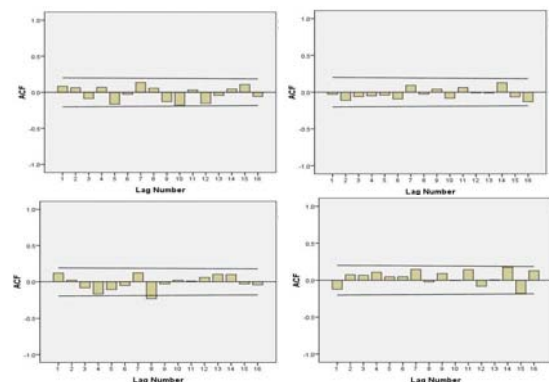


ภาพที่ 3 SACF และ SPACF ของ SO₂, NO₂, CO และ
O₃ ที่สแตนดาร์แล้ว เรียงจากบนลงล่าง

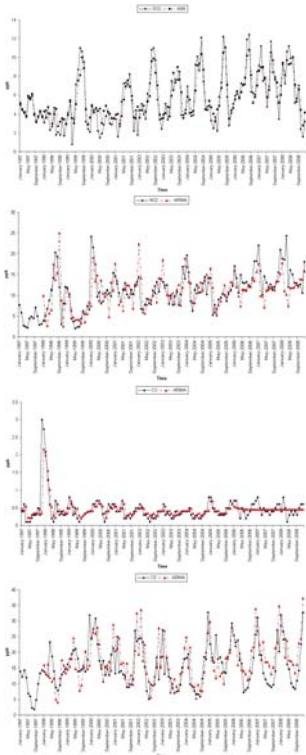
2) ประมาณค่าตัวแบบที่กำหนด ด้วยวิธีกำลัง
สองน้อยที่สุดไม่เชิงเส้น ใช้โปรแกรม SPSS ช่วยในการ
ประมาณค่า

3) ตรวจสอบความเหมาะสม พิจารณาจากกราฟ
ACF ของ residual ดังภาพที่ 4 พบว่า ทุกตัวแบบมีความ
เหมาะสม และที่ CO จะพบว่า มี residual ที่ lag ที่ 8 เกินช่วง
เส้นประขนานคู่ ที่ 95% ช่วงความเชื่อมั่น (confident band)
ค่า residual ที่ lag 8 นี้ เป็นค่าสังเกตนอกกรอบ (outlier)
ต้องใช้ทฤษฎีเรื่อง ARIMA intervention ช่วย ซึ่งใน
งานวิจัยนี้ไม่ขอกล่าวถึง

4) นำตัวแบบที่ได้มาพยากรณ์กับชุดข้อมูล
ทดสอบได้แก่ พยากรณ์ระยะสั้น (3 เดือน) ใช้ข้อมูลเดือนที่
109-111 และพยากรณ์ระยะยาว (36 เดือน) ใช้ข้อมูลเดือนที่
109-144 ดังภาพที่ 5 และได้ค่าความคลาดเคลื่อนดังตาราง
ที่ 1



ภาพที่ 4 ACF ของ residual SO₂, NO₂, CO และ O₃ เรียง
จากซ้ายไปขวาและบนลงล่าง



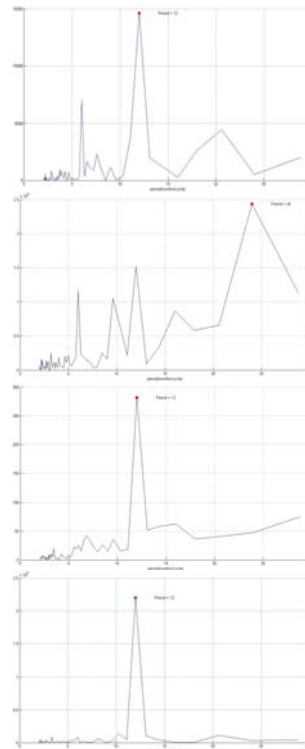
ภาพที่ 5 ตัวแบบ ARIMA ใช้พยากรณ์ข้อมูลชุด

ทดสอบ SO₂, NO₂, CO และ O₃ เรียงจากซ้ายไปขวาและบนลงล่าง

4. ตัวแบบ ANN นำข้อมูลมาแปลงค่าด้วยวิธีการนอร์มัลไลซ์ข้อมูล (Normalization) เป็นการลดค่าของข้อมูลให้อยู่ในขอบเขตที่น้อยลงเพื่อให้เหมาะสมกับฟังก์ชันที่ใช้ในโครงข่ายประสาทเทียม (ฟังก์ชัน sigmoid) คือช่วง 0-1 แล้วแบ่งข้อมูลออกเป็น 2 ชุด ได้แก่ ชุดข้อมูลสำหรับหาตัวแบบจำนวน 108 ข้อมูล และชุดข้อมูลทดสอบ 36 ข้อมูล ในชุดข้อมูลหาตัวแบบแบ่งข้อมูลออกเป็นชุดข้อมูลทดสอบ 36 ข้อมูล ชุดข้อมูลเรียนรู้ 50% ที่เหลือเป็นชุดข้อมูลสำหรับ validation และเมื่อได้ตัวแบบแล้วนำมาพยากรณ์กับชุดข้อมูลทดสอบ พยากรณ์ระยะสั้น (3 เดือน) ใช้ข้อมูลเดือนที่ 109-111 และพยากรณ์ระยะยาว (36 เดือน) ใช้ข้อมูลเดือนที่ 109-144 และจำนวนชั้นซ่อนในงานวิจัยนี้ใช้เพียงชั้นเดียวก็เพียงพอต่อข้อมูลที่ใช้และการประมวลผลไม่ซับซ้อนยุ่งยาก

1) การหาค่า input node ใช้แบล็กบ็อก Fast Fourier Transform (fft) ช่วยในการหาคาบของข้อมูล

อนุกรมเวลาใน matlab ดังภาพที่ 6 จากผลการทดลองได้คาบของ SO₂ เท่ากับทุก 24 เดือน NO₂ เท่ากับทุก 12 เดือน CO เท่ากับทุก 12 เดือน และ O₃ เท่ากับทุก 12 เดือน นำไปใช้กำหนด input node

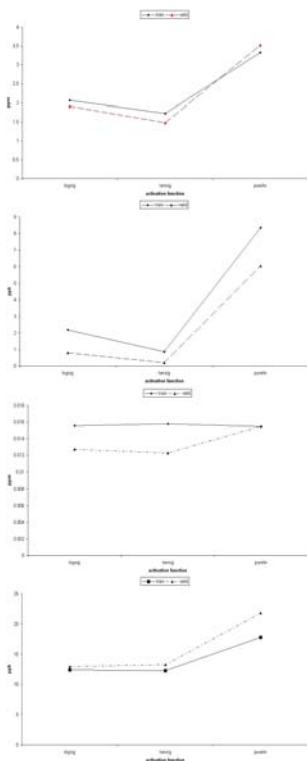


ภาพที่ 6 ใช้แบล็กบ็อก fft หาคาบของ SO₂, NO₂, CO และ O₃ เพื่อใช้กำหนด input node เรียงลำดับจากซ้ายไปขวาและบนลงล่าง

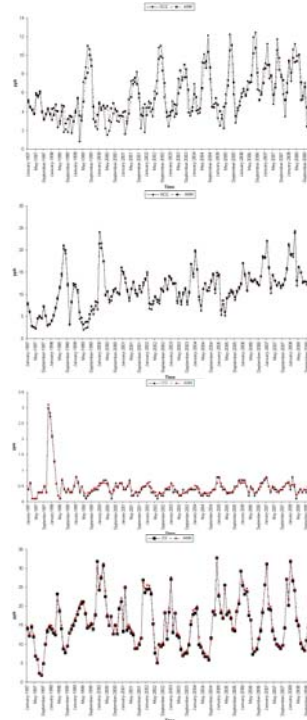
2) หาจำนวน hidden node ที่เหมาะสม ใช้ input node จากข้อที่ 4.1 กำหนดจำนวนรอบเท่ากับ 5 อัตราการเรียนรู้ 0.01 ฟังก์ชันกระตุ้นใช้ Log-sigmoid ฟังก์ชันเทรนนิ่งใช้ trainlm ฟังก์ชันเรียนรู้ใช้ learnsgdm ค่าความคลาดเคลื่อน (MSE) เป้าหมาย 0.00005 ทำการทดลองซ้ำ 30 ครั้ง และหาค่าความคลาดเคลื่อนเฉลี่ย จากชุดข้อมูลทดสอบในชุดข้อมูลสำหรับหาตัวแบบกับค่าพยากรณ์ พบว่าเมื่อเพิ่มจำนวน hidden node มากขึ้น ค่าความคลาดเคลื่อนลดลงเรื่อย ๆ จึงไม่สามารถเลือกจำนวน hidden node จากการทดลองได้ จึงใช้ทฤษฎีวิธีการหาค่าเฉลี่ยของผลบวกระหว่างจำนวนโหนดชั้นข้อมูลเข้าชั้นและจำนวนโหนดชั้นข้อมูลออก (Berke and Hajela, 1991) เพื่อกำหนด hidden

node แทน เพราะฉะนั้น SO_2 , NO_2 , CO และ O_3 มีจำนวน hidden node เท่ากับ 7, 13, 7 และ 7 ตามลำดับ

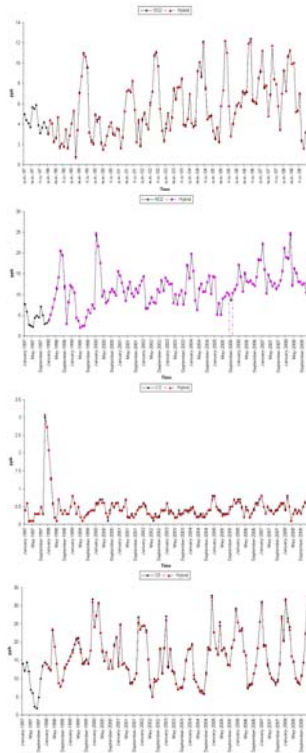
3) หาฟังก์ชันกระตุ้นที่เหมาะสม ได้แก่ Log-sigmoid, Hyperbolic tangent sigmoid และ Linear transfer ใช้ข้อมูล input node และ hidden node จากข้อที่ 4.1 และ 4.2 กำหนดอัตราการเรียนรู้ 0.01 และจำนวนรอบเท่ากับ 10 ฟังก์ชันเทรนนิ่งใช้ trainlm ฟังก์ชันเรียนรู้ใช้ learnqdm ค่าความคลาดเคลื่อน (MSE) เป้าหมาย 0.00005 ทำการทดลองซ้ำ 30 ครั้งและหาค่าความคลาดเคลื่อน MSE เฉลี่ย ดังภาพที่ 7 พบว่า SO_2 , NO_2 และ CO ฟังก์ชันกระตุ้นที่เหมาะสมคือ hyperbolic tangent sigmoid ส่วนก๊าซ O_3 ฟังก์ชันกระตุ้นที่เหมาะสมคือ log-sigmoid



ภาพที่ 7 เปรียบเทียบค่าความคลาดเคลื่อน MSE ที่ได้จากการหาฟังก์ชันกระตุ้นของ SO_2 , NO_2 , CO และ O_3 เรียงจากซ้ายไปขวาและบนลงล่าง

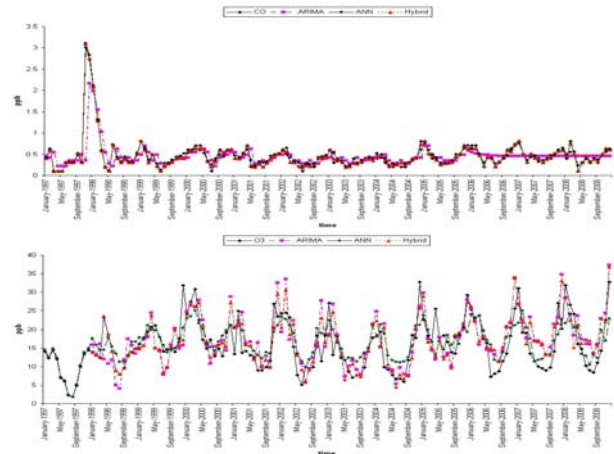


ภาพที่ 8 ตัวแบบ ANN พยากรณ์ข้อมูลชุดทดสอบ SO_2 , NO_2 , CO และ O_3 เรียงจากซ้ายไปขวาและบนลงล่าง



ภาพที่ 9 ตัวแบบผสม พยากรณ์ข้อมูลชุดทดสอบ

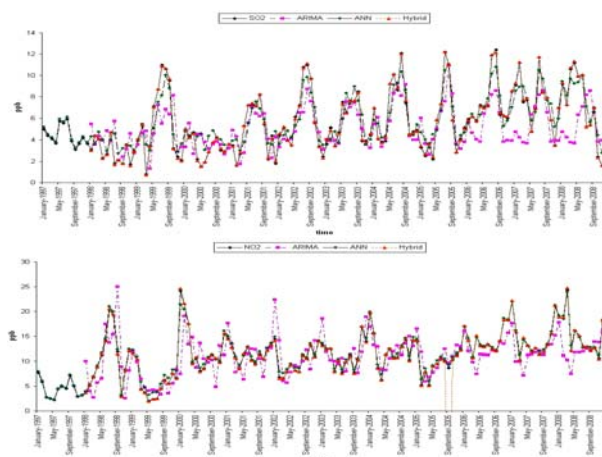
SO₂, NO₂, CO และ O₃ เรียงจากซ้ายไปขวา
และบนลงล่าง



ภาพที่ 10 เปรียบเทียบการพยากรณ์ทั้งสามตัวแบบกับชุด

ข้อมูลทดสอบ SO₂, NO₂, CO และ O₃ เรียงจาก
บนลงล่าง

4) นำตัวแบบที่ได้พยากรณ์ข้อมูลทดสอบ
กำหนดจำนวนรอบเท่ากับ 50 อัตราการเรียนรู้ 0.01
ฟังก์ชันเทรนนิ่ง



ตารางที่ 1 เปรียบเทียบค่าความคลาดเคลื่อนด้วยแบบที่ใช้พยากรณ์ก๊าซแต่ละชนิด

มลภาวะ ใน อากาศ	ตัวแบบ	MSE			MAPE		
		train(108)	test(3)	test(36)	train(108)	test(3)	test(36)
SO ₂	ARIMA(1,0,0)(0,1,1) ₁₂	3.6757	1.7433	10.680	42.1291	19.7417	37.0907
	ANN 12-7-1 (tansig)	1.4692	0.0849	1.3336	29.4375	4.7207	16.8989
	ตัวแบบผสม	0.0131	0.0045	0.0102	2.3109	1.1115	1.1105
NO ₂	ARIMA(1,0,0)(1,1,1) ₁₂	14.2814	2.0994	15.9148	31.3704	9.71458	16.9073
	ANN 24-13-1 (tansig)	0.4058	0.0720	0.1447	7.0338	1.6266	2.1982
	ตัวแบบผสม	19.5688	0.0027	0.0164	6.2289	0.3508	0.5254
CO	ARIMA(1,0,0)(0,0,0) ₁₂	0.0928	0.0183	0.0262	44.0924	15.7143	40.2202
	ANN 12-7-1 (tansig)	0.0047	0.0052	0.0053	16.6405	8.6663	15.7024
	ตัวแบบผสม	0.0012	0.0025	0.0009	7.9942	7.3049	5.5256
O ₃	ARIMA(0,0,3)(1,1,0) ₁₂	29.9538	21.3942	34.0250	29.1552	12.9361	37.6690
	ANN 12-7-1 (logsig)	0.6500	0.9682	0.4838	4.8688	4.0679	3.9311
	ตัวแบบผสม	0.1357	0.1094	0.1319	1.7607	1.0534	2.1954

ตารางที่ 2 ตัวแบบที่ให้ค่าการพยากรณ์แม่นยำที่สุดของก๊าซแต่ละชนิด

มลภาวะ ใน อากาศ	ตัวแบบที่ให้ค่าการพยากรณ์ที่แม่นยำที่สุด	
	ระยะสั้น (3 เดือน)	ระยะยาว (36 เดือน)
SO ₂	ตัวแบบผสม ARIMA(1,0,0)(0,1,1) ₁₂ รวมกับ ANN 12-7-1 (tansig)	ตัวแบบผสม ARIMA(1,0,0)(0,1,1) ₁₂ รวมกับ ANN 12-7-1 (tansig)
NO ₂	ตัวแบบผสม ARIMA(1,0,0)(1,1,1) ₁₂ รวมกับ ANN 24-13-1 (tansig)	ตัวแบบผสม ARIMA(1,0,0)(1,1,1) ₁₂ รวมกับ ANN 24-13-1 (tansig)
CO	ตัวแบบผสม ARIMA(1,0,0)(0,0,0) ₁₂ รวมกับ ANN 12-7-1 (tansig)	ตัวแบบผสม ARIMA(1,0,0)(0,0,0) ₁₂ รวมกับ ANN 12-7-1 (tansig)
O ₃	ตัวแบบผสม ARIMA(0,0,3)(1,1,0) ₁₂ รวมกับ ANN 12-7-1 (logsig)	ตัวแบบผสม ARIMA(0,0,3)(1,1,0) ₁₂ รวมกับ ANN 12-7-1 (logsig)

ใช้ trainlm ฟังก์ชันเรียนรู้ใช้ learnqdm และค่าความคลาดเคลื่อนเป้าหมาย 0.00005 ทำการทดลองซ้ำ 30 ครั้งและหาค่าความคลาดเคลื่อนเฉลี่ย นำตัวแบบมาพยากรณ์ดังภาพที่ 8 และค่าความคลาดเคลื่อนจากการพยากรณ์แสดงในตารางที่ 1

5) ตัวแบบผสม ใช้ตัวแบบ ARIMA พยากรณ์ข้อมูลจริง จะได้ค่าความคลาดเคลื่อนหรือค่า residual และใช้ตัวแบบ ANN พยากรณ์ค่า residual แล้วนำค่าพยากรณ์จากทั้งสองตัวแบบมารวมกันดังแสดงในภาพที่ 9 และได้ค่าความคลาดเคลื่อนจากการพยากรณ์ในตารางที่ 1

ผลการทดลอง

ผลการศึกษาตัวแบบผสมระหว่างวิธีบ็อกซ์และเจนกินส์กับวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ สำหรับก๊าซซัลเฟอร์ไดออกไซด์ (SO₂) ก๊าซไนโตรเจนไดออกไซด์ (NO₂) ก๊าซคาร์บอนมอนอกไซด์ (NO) และก๊าซโอโซน (O₃) ของสถานีตรวจวัดคุณภาพอากาศอนามัยมาตาปุด ตำบลมาตาปุด อำเภอเมือง จังหวัดระยอง (เขตนิคมอุตสาหกรรมมาตาปุด) เมื่อเปรียบเทียบค่าความคลาดเคลื่อน MSE และ MAPE ดังแสดงในตารางที่ 1 และสรุปออกมาเป็นตัวแบบที่ให้ค่าพยากรณ์แม่นยำที่สุดในตารางที่ 2 พบว่า ตัวแบบผสมจะให้ผลการพยากรณ์ที่แม่นยำมากกว่าตัวแบบ ANN และตัวแบบ ARIMA และตัวแบบ ARIMA จะให้การพยากรณ์ที่แม่นยำน้อยที่สุดทั้งการพยากรณ์ระยะสั้น (3 เดือน) และพยากรณ์ระยะยาว (36 เดือน)

สรุปและข้อเสนอแนะ

งานวิจัยในครั้งนี้ เนื่องจากข้อมูลที่ใช้ในการพยากรณ์เป็นข้อมูลเฉลี่ยรายเดือน แต่ในสภาพแวดล้อมจริงนั้นมีปัจจัยที่ทำให้คุณภาพอากาศเปลี่ยนแปลงนั้นหลายปัจจัย ไม่ว่าจะเป็นอุณหภูมิ ทิศทางลม ฤดูกาล กระแสน้ำร้อนน้ำเย็น สนามแม่เหล็กโลก เพื่อที่จะทำการพยากรณ์แม่นยำมากยิ่งขึ้น ควรนำปัจจัยเหล่านี้

เข้ามาเป็นพารามิเตอร์ประกอบ และข้อมูลควรจะมีจำนวนมากว่านี้ และเนื่องจากวิธีการทดลองอาศัยเครื่องมือหลายชนิดที่มีอยู่ทั่วไป (off-the-shelf s/w tools) จึงควรนำมาประกอบกันจนเป็นเครื่องมือที่นำมาใช้งานได้จริง

เอกสารอ้างอิง

- ทรงศิริ แด่สมบัติ. 2549. เทคนิคการพยากรณ์เชิงปริมาณ. กรุงเทพฯ : ภาควิชาสถิติ มหาวิทยาลัยเกษตรศาสตร์.
- พินดา ผดุงเวียง, สิริภัทร เชื้อขวชาญวัฒนา และคำธณัฐติ. 2009. ระบบตัวช่วยแนะแนวการเลือกกลุ่มคณะเข้าเรียนต่อในมหาวิทยาลัย. Proc. Conf. on Knowledge and Smart Technology.7(24-25): 89-95.
- นันทวัฒน์ อินทร์รุ่งเรือง. 2551. ภาวะมลพิษทางอากาศในประเทศไทย. ค้นเมื่อ 3 มกราคม 2552, จาก http://www.reo11.net/download/news/Pullution_of_Air_in_Thailand.pdf
- หนังสือพิมพ์มติชนออนไลน์. 2551. สถานการณ์มลพิษในประเทศไทย ปี 2551. ค้นเมื่อ 3 มกราคม 2552, จาก http://www.matichon.co.th/news_detail.php?newsid=1230543793&grpId=01&catid=042552
- Asha B. Chelani and S. Devotta. 2006. Air quality forecasting using a hybrid autoregressive and nonlinear model. Atmospheric Environment. 10 (40): 1774-1780.
- Behzad Hashemloo. 2008. Forecasting pavement surface temperature using time series and artificial neural networks. Retrieved October 10, 2009, from <http://uwspace.uwaterloo.ca/bitstream/10012/3797/1/My%20Thesis5.pdf>

- Berke L. and Hajela P. 1991. Application of neural Networks in structural optimization. NATO/AGARD Advanced Study Institute, 23(I-II): 731-745.
- G. Peter Zhang. 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing. 1(50): 159-175.
- Giuseppe Nunnari and Flavio Cannav'. 2005. Modified cost functions for modeling air quality time series by using neural networks. LNCS. 3697: 723-728.
- Luis A. Diaz-Robles et al. 2008. A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. Atmosphere Environment. 35(42): 8331-8340.
- The Mathworks Inc. 2007. Matlab: Matlab7 data analysis. MA, USA.