

การพยากรณ์ความต้องการน้ำของพืชอ้างอิงที่มีค่าสูญหาย
ในประเทศไทย

PREDICTION OF REFERENCE EVAPOTRANSPIRATION WITH
MISSING DATA IN THAILAND



T139323



ณ.
๐๘๗/ก
๒๕๕๖



b.12๕211๙๐

เลขหมู่.....
เลขทะเบียน..... 139323
วันเดือนปี 130 ต.ค. 2558

รายงานนี้เป็นส่วนหนึ่งของ การศึกษาอิสระ 2
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาคเรียนที่ 2 ปีการศึกษา 2556

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**PREDICTION OF REFERENCE EVAPOTRANSPIRATION WITH
MISSING DATA IN THAILAND**



**A REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
OF THE COURSE**

INDEPENDENT STUDY 2

MASTER OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2/ 2013

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2014

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ	การพยากรณ์ความต้องการน้ำของพืชอ้างอิงจากข้อมูลสภาพอากาศที่มีค่าสูญหายในประเทศไทย
นักศึกษา	นายเอก ธรรมวิวัฒน์
รหัสนักศึกษา	54660502
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
แขนงวิชา	เทคโนโลยีระบบสารสนเทศ
ปีการศึกษา	2556
อาจารย์ที่ปรึกษา	ดร. กิติ์สุชาติ พสุภา

บทคัดย่อ

โครงข่ายประสาทเทียม (Artificial Neural Network; ANN) ได้ถูกนำมาใช้ในการหาความต้องการน้ำอ้างอิงของพืช โดยมีประสิทธิภาพใกล้เคียงกับวิธีที่ใช้กันอย่างแพร่หลายคือวิธี “Penman-Monteith” ในการศึกษาพบว่าเราสามารถหาความต้องการน้ำอ้างอิงของพืชได้การใช้โครงข่ายประสาทเทียมกับข้อมูลสภาพอากาศในประเทศไทย มีประสิทธิภาพใกล้เคียงกับวิธี Penman-Monteith ในการศึกษานี้ได้เพิ่มปัญหาการสูญหายของข้อมูลสภาพอากาศ กรณีที่มีข้อมูลสูญหายนั้นเราสามารถใช้อัลกอริทึม Expectation-Maximization เพื่อแก้ปัญหาข้อมูลสูญหายได้ โดยความแม่นยำในการหาความต้องการน้ำของพืชอ้างอิง จะแปรผกผันกับจำนวนค่าสูญหายของสภาพอากาศ ได้มีการทดลองศึกษาตัวแปรลักษณะจำเพาะของสภาพอากาศ แสดงให้เห็นถึงตัวแปรลักษณะจำเพาะของสภาพอากาศที่สำคัญคือ ความยาวนานแสงแดด เป็นตัวแปรที่มีความสำคัญที่สุดตามมาด้วย อุณหภูมิ และ ความเร็วลม ตามลำดับ

Title Prediction of Reference Evapotranspiration with Missing Data in Thailand

Student Ek Thamwiwatthana

Student ID. 54660502

Degree Master of Science

Program Information Technology

Major Information System Technology

Academic Year 2013

Advisor Dr. Kitsuchart Pasupa

ABSTRACT

Artificial Neural Networks (ANNs) has been used in prediction of reference evapotranspiration for a recent decade. Its performance is competitive to a widely used method they so-called “Penman-Monteith” method. In this study, we aim to estimate the crop evapotranspiration by ANNs from climatic data in Thailand and compare the performance with the Penman-Monteith method. As missing data is inevitable, we also included the missing data situation into the study. This can be solved by expectation-maximization algorithm. The accuracy of the prediction decreases when the amount of missing values increases. Furthermore, we exploit the feature selection in the study. It shows that sunshine duration is the most important feature followed by temperature and wide speed, respectively.

กิตติกรรมประกาศ

การศึกษาระดับนี้สำเร็จได้ด้วยความกรุณาของ ดร. กิติ์สุชาติ พสุภา ซึ่งเป็นอาจารย์ที่ปรึกษา สำหรับความช่วยเหลือ ให้คำชี้แนะ ช่วยแก้ปัญหาตลอดจนให้ความรู้และประสบการณ์ที่ดีแก่ข้าพเจ้า ขอขอบคุณ ดร. สิงหะ นวีสุข สำหรับคำแนะนำ และความช่วยเหลือต่างๆ โดยเฉพาะสถานที่ห้องปฏิบัติการ ขอขอบคุณ พี่ๆ น้องๆ ใน Multidisciplinary Lab ทุกคนสำหรับความช่วยเหลือ และกำลังใจที่ดีต่อข้าพเจ้าตลอดมา ขอขอบคุณการสนับสนุนการทำการศึกษาระยะจากสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ขอขอบคุณกรมอุตุนิยมวิทยาสำหรับข้อมูลสภาพอากาศที่ใช้ในการทำวิจัย สุดท้ายนี้ขอขอบคุณครอบครัวและเพื่อนๆ ของข้าพเจ้า สำหรับการสนับสนุน การส่งเสริมในทุกๆ ด้านในงานวิจัย รวมถึงกำลังใจที่ได้ตลอดเวลาที่ทำงานวิจัย จนสำเร็จลุล่วงเป็นอย่างดี คุณงามความดีอันใดที่เกิดจากการศึกษาระดับนี้ ข้าพเจ้าขอมอบให้กับบิดามารดา ซึ่งเป็นที่รักและเคารพยิ่ง ตลอดจนครูบาอาจารย์ที่เคารพทุกท่านที่ได้มอบวิชาความรู้และถ่ายทอดประสบการณ์ที่ดีให้แก่ข้าพเจ้า

เอก ธรรมวิวัฒน์

สารบัญ

	หน้า
บทคัดย่อ.....	I
บทคัดย่อภาษาอังกฤษ	II
กิตติกรรมประกาศ.....	III
สารบัญ	IV
สารบัญตาราง	VI
สารบัญรูป	VII
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์.....	3
1.3 ขอบเขตงานวิจัย	3
1.4 วิธีการดำเนินงาน.....	3
1.5 ประโยชน์ที่คาดว่าจะได้รับ	4
1.6 ผลลัพธ์จากการศึกษา.....	4
บทที่ 2 ทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้อง	5
2.1 งานวิจัยที่เกี่ยวข้อง.....	5
2.2 ความต้องการน้ำของพืช	6
2.3 โมเดลที่ใช้ในการคำนวณหาค่าการคายน้ำอ้างอิงของพืช.....	8
2.4 การเรียนรู้ของเครื่อง (Machine Learning).....	8
2.5 ระบบโครงข่ายประสาทเทียม (Artificial Neural Network).....	10
2.6 Back-Propagation Algorithm	12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และ IV อ่างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
2.7 ความถดถอยเชิงเส้น (Linear Regression)	13
2.7 การแก้ปัญหาค่าสูญหายของข้อมูล	14
2.8 Expectation Maximization Algorithm	14
บทที่ 3 การออกแบบการทดลอง	16
บทที่ 4 ผลการทดลอง.....	20
4.1 การหาความต้องการน้ำอ้างอิงของพืชโดยที่ไม่มีค่าสูญหายของสภาพภูมิอากาศ	20
4.2 การหาความต้องการน้ำของพืชอ้างอิงโดยมีค่าสูญหายของสภาพภูมิอากาศ	22
4.3 การหาความต้องการน้ำของพืชอ้างอิงโดยการเลือกใช้ตัวแปรลักษณะจำเพาะ	26
บทที่ 5 สรุปและวิจารณ์ผลการทดลอง	32
บรรณานุกรม	33
ภาคผนวก	35
บทความวิจัยที่ได้เผยแพร่และตีพิมพ์.....	36

สารบัญตาราง

ตารางที่	หน้า
3.1 แสดงตัวแปรลักษณะจำเพาะของข้อมูลสภาพอากาศที่ใช้ในการหาความต้องการน้ำอ้างอิงของพืช.....	17
4.1 แสดงค่าเฉลี่ยผลลัพธ์การหาความต้องการน้ำอ้างอิงของพืชโดยใช้วิธีการคำนวณแบบ ANN และ LR โดยที่ไม่มีค่าสูญหาย	20
4.2 แสดงค่าเฉลี่ยผลลัพธ์ของการหาความต้องการน้ำอ้างอิงของพืชโดยมีค่าสูญหาย ที่มีการทดแทนค่าสูญหายด้วยวิธี EM เปรียบเทียบระหว่าง วิธีการคำนวณแบบ ANN และ LR	23
4.3 ผลการเปรียบเทียบประสิทธิภาพระหว่างการหาความต้องการน้ำอ้างอิงของพืชด้วยวิธี ANN และ LR เมื่อเลือกใช้ตัวแปรลักษณะจำเพาะ	24
4.4 ผลการเปรียบเทียบประสิทธิภาพระหว่างการหาความต้องการน้ำอ้างอิงของพืชด้วยวิธี ANN และ LR เมื่อเลือกใช้ตัวแปรลักษณะจำเพาะ และมีค่าสูญหาย	28
4.5 ผลการเปรียบเทียบประสิทธิภาพระหว่างการหาความต้องการน้ำอ้างอิงของพืชด้วยวิธี ANN เมื่อเลือกใช้ตัวแปรลักษณะจำเพาะ และมีค่าสูญหาย	30

สารบัญรูป

รูปที่	หน้า
2.1 ระบบเครือข่ายประสาท.....	10
2.2 รูปแบบการคำนวณของ Neural Network	11
2.3 แสดงลักษณะของ Sigmoid Function	12
2.4 แสดงโครงสร้างของ โครงข่ายประสาทเทียม.....	12
3.1 แสดงแผนที่จังหวัดเชียงใหม่ และจังหวัดอุบลราชธานี	16
3.2 แสดงภาพการทำงานของ 5-fold Cross-Validation	18
3.3 แสดงภาพรวมของการทำงานของวิธี Machine Learning	19
4.1 แสดงการเปรียบเทียบของผลลัพธ์ความต้องการน้ำอ้างอิงของพีชระหว่างวิธีการคำนวณแบบ ANN และวิธี PM จากข้อมูลของสถานีตรวจอากาศจังหวัดเชียงใหม่ในปี พ.ศ. 2549-2554	21
4.2 แสดงการเปรียบเทียบของผลลัพธ์ความต้องการน้ำอ้างอิงของพีชระหว่างวิธีการคำนวณแบบ ANN และวิธี PM จากข้อมูลของสถานีตรวจอากาศจังหวัดอุบลราชธานีในปี พ.ศ. 2549-2554.....	22
4.3 แสดงค่าเฉลี่ยผลลัพธ์ของการหาความต้องการน้ำอ้างอิงของพีชโดยมีค่าสูญหาย ที่มีการทดแทนค่าสูญหายด้วยวิธี EM เปรียบเทียบระหว่าง วิธีการคำนวณแบบ ANN และ LR	23
4.4 แสดงค่าเฉลี่ยผลลัพธ์ของการหาความต้องการน้ำอ้างอิงของพีชโดยมีค่าสูญหาย ที่มีการทดแทนค่าสูญหายด้วยวิธี EM และ การทดแทนค่าสูญหายจากวันที่ใกล้ที่สุด โดยใช้วิธี ANN ในการคำนวณ	26
4.5 แสดงค่าเฉลี่ย ความสำคัญของตัวแปรลักษณะจำเพาะ ในรูปแบบของค่าน้ำหนัก ($ w_j $) LR ของทุกการทดลอง LR ในการทดลองที่ 1	27
4.6 แสดงความสัมพันธ์ของการเพิ่มขึ้น/ลดลง ของความแม่นยำของการหาความต้องการน้ำอ้างอิงของพีชด้วยวิธี ANN และ LR	29

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

อาชีพเกษตรกรนั้นเป็นอาชีพของคนส่วนใหญ่ในประเทศไทย ประเทศไทยมีประชากรจำนวน 63 ล้านคน ซึ่งกว่า 40 ล้านคนหรือประมาณร้อยละ 63 มีอาชีพทางการเกษตร ซึ่งมีพื้นที่ถือครองทางการเกษตรเป็นพื้นที่กว่าร้อยละ 41 และมีแนวโน้มว่าจะขยายตัวเพิ่มขึ้น [1] การขาดแคลนน้ำในฤดูแล้งยังพบได้ทั่วไปทุกภูมิภาคในประเทศไทย เนื่องจากเราสามารถสร้างระบบกักเก็บน้ำได้เพียงร้อยละ 33.5 ของความต้องการการใช้น้ำทั้งหมด เท่านั้น และยังไม่สามารถที่จะนำมาใช้งานจริงได้ทั้งหมด ซึ่งไม่เพียงพอในการใช้งานในหน้าแล้ง ทำให้ไม่สามารถสร้างผลผลิตทางการเกษตรได้อย่างมีประสิทธิภาพ ดังนั้นการประยุกต์ใช้ทรัพยากรเทคโนโลยีสารสนเทศเข้ามาจัดการและช่วยเหลือทางด้านเกษตรจึงเป็นเรื่องที่จำเป็นจะต้องเกิดขึ้น เพื่อให้เกิดการลดต้นทุนประหยัดทรัพยากร และเหมาะสมกับแต่ละพื้นที่ ซึ่งจะทำให้เกิดการใช้ทรัพยากรอย่างคุ้มค่า ดังนั้นเทคโนโลยีสารสนเทศสามารถนำมาใช้กับการเกษตรได้ ยกตัวอย่างเช่นประเทศอินเดียและญี่ปุ่นนำเทคโนโลยีสารสนเทศเข้ามาประยุกต์ใช้กันอย่างกว้างขวางเพื่อเพิ่มความสามารถในการเพิ่มผลผลิตทางการเกษตรและแก้ปัญหาการขาดแคลนทรัพยากร ดังนั้นเราควรให้ความสำคัญกับการใช้เทคโนโลยีสารสนเทศมาพัฒนาการเกษตร เนื่องจากประเทศไทยเป็นประเทศเกษตรกรรม จึงต้องเริ่มให้ความสนใจด้านการประยุกต์ใช้เทคโนโลยีสารสนเทศกับการเกษตรเพื่อไม่ให้เสียโอกาสในการแข่งขันกับประเทศต่างๆ และช่วยให้การเก็บข้อมูลเกี่ยวกับการเกษตรสามารถทำได้อย่างมีประสิทธิภาพมากยิ่งขึ้น เพื่อได้ข้อมูลที่สามารถนำมาประมวลผลต่อไป

การพยากรณ์หาการคายน้ำของพืชอ้างอิง เป็นสิ่งหนึ่งที่จะช่วยให้การจัดการน้ำเป็นไปอย่างมีประสิทธิภาพ ศักยภาพความต้องการน้ำของพืช คือ ผลรวมของการสูญเสียน้ำที่พืชสูญเสียน้ำไปในอากาศประกอบด้วย น้ำจากการคายน้ำของพืช และการระเหยของน้ำจากผิวดิน ที่สามารถหาได้จากวิธีการวัดจาก Lysimeter ซึ่งเป็นวิธีการที่ใช้เวลา และต้องการความแม่นยำในการทำการทดลอง อีกวิธีคือการวัดจากสภาพอากาศของบริเวณเพาะปลูก ในการคำนวณจะใช้ วิธีการ Penman-Monteith (PM) ซึ่งมีประสิทธิภาพและถูกแนะนำให้เป็นวิธีมาตรฐาน เพื่อหาค่าความต้องการน้ำของพืชอ้างอิง (Reference Evapotranspiration; ET_0) [2] เป็นวิธีที่มีการคำนวณซับซ้อนในการคำนวณหาค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความต้องการน้ำของพืช เนื่องจากใช้ข้อมูลของสภาพอากาศและสิ่งแวดล้อมหลายตัวแปรในการคำนวณ และไม่สามารถทำการคำนวณได้เมื่อมีค่าสูญหายเกิดขึ้น มีการศึกษาหาวิธีการพยากรณ์ความต้องการน้ำของพืชหลายแบบ วิธีโครงข่ายประสาทเทียม (ANN) ก็เป็นวิธีที่ถูกนำมาใช้เพื่อพยากรณ์หาค่าความต้องการน้ำของพืช ANN จะถูกนำมาใช้สอนโครงข่ายและนำมาทดสอบกับข้อมูลสภาพอากาศ เพื่อหาผลลัพธ์ ค่าความต้องการน้ำของพืช ซึ่งจะได้ผลลัพธ์ที่มีประสิทธิภาพสูง มีงานวิจัยที่แสดงผลว่าการใช้โครงข่ายประสาทเทียมแบบ Single Hidden Neuron นั้นเพียงพอที่จะใช้สำหรับการคำนวณเพื่อหาค่า ET_0 เมื่อใช้วิธีการ PM ซึ่งมีข้อมูลสภาพอากาศประจำวัน คือ อุณหภูมิสูงสุดและต่ำสุด ความชื้นสูงสุดและต่ำสุด ความเร็วลมและการแผ่รังสีของดวงอาทิตย์ (Solar Radiation) [2] นอกจากนี้ยังมีการศึกษาพบว่า ANN จะมีการคำนวณหา การระเหยของน้ำ (Evaporation) ได้มีประสิทธิภาพดี เมื่อมีการใช้ข้อมูลผสม ระหว่าง ค่าเฉลี่ยของอุณหภูมิอากาศ ความเร็วลม ชั่วโมงแสงแดด และความชื้น [3] นอกจากนี้ มีการทดลองเพื่อทดสอบ โดยมีผลลัพธ์ว่าเมื่อใช้ค่าอุณหภูมิของอากาศ ความเร็วลม ความชื้น และ การแผ่รังสีของดวงอาทิตย์ จะให้ผลลัพธ์ที่ดีที่สุด ซึ่งในการทดลอง การเพิ่มตัวแปรมากขึ้นจะให้ผลลัพธ์ที่ดียิ่งขึ้น แต่ในบางพื้นที่ ข้อมูลสภาพอากาศถูกจำกัด ข้อมูลอุณหภูมิของอากาศและการแผ่รังสีดวงอาทิตย์ก็สามารถหาการระเหยน้ำได้ [4] ในการเก็บข้อมูลสภาพอากาศ ต้องใช้อุปกรณ์หลายชนิดเพื่อเก็บค่าตัวแปรเพื่อใช้ในการคำนวณ ซึ่งในบางพื้นที่ไม่สามารถที่จะทำได้อาจจะด้วยงบประมาณ การขาดแคลนอุปกรณ์ หรือข้อจำกัดอื่นๆ ทำให้ไม่สามารถจัดเตรียมอุปกรณ์ได้ครบ จึงไม่สามารถคำนวณด้วยวิธีการปกติได้ ดังนั้นการลดจำนวนตัวแปรเป็นสิ่งหนึ่งที่สำคัญ นอกจากจะช่วยในเรื่องของลดจำนวนของอุปกรณ์แล้ว ยังสามารถช่วยในการลดข้อจำกัดในการคำนวณอีกด้วย ข้อมูลสภาพอากาศที่นำมาใช้ ในการทำงานจริงอาจจะมีข้อมูลสูญหายเกิดขึ้นได้ ซึ่งเกิดขึ้นได้จากสาเหตุหลายประการ อาจจะเป็นปัญหาจากการจัดเก็บข้อมูลหรือ ปัญหาจากการตรวจวัดสภาพอากาศ เช่น อุปกรณ์ไม่สามารถอ่านค่าได้ หรืออุปกรณ์มีความเสียหายชั่วคราว ทำให้ไม่สามารถทำการคำนวณหาความต้องการน้ำของพืชได้ หรืออาจใช้วิธีอื่นเข้ามาช่วยเช่น การนำค่าเฉลี่ยมาใช้ แต่อาจทำให้ความแม่นยำในการคำนวณลดลง ในกรณีที่มีจำนวนของค่าที่สูญหายเป็นจำนวนมาก นับได้ว่าเป็นปัญหาอย่างหนึ่งที่จะเกิดขึ้นได้ เมื่อนำการหาความต้องการน้ำของพืชไปใช้ในทางปฏิบัติ

จุดประสงค์หลักของการศึกษาเพื่อพัฒนาวิธีการพยากรณ์หาความต้องการน้ำของพืช โดยใช้วิธี ANN เปรียบเทียบประสิทธิภาพในการพยากรณ์ความต้องการน้ำของพืชกับวิธีการหาความต้องการน้ำของพืชแบบอื่นๆ อีกทั้งเพื่อลดจำนวนตัวแปรสภาพอากาศที่ใช้ในการคำนวณ หาค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ET_0 โดยใช้วิธีการ Linear Regression (LR) ซึ่งใช้ในการวิเคราะห์ เพื่อจัดลำดับประสิทธิภาพของตัวแปรสภาพอากาศที่จะนำมาใช้เพื่อลดตัวแปรในการคำนวณ แล้วนำ Expectation-Maximization Algorithm (EM) และวิธีการ k-Nearest Neighbor มาใช้เพื่อแก้ปัญหาข้อมูลสูญหาย ช่วยลดความผิดพลาดของข้อมูล ซึ่งจะทำให้วิธีการพยากรณ์หาค่าความต้องการน้ำของพืชมีความยืดหยุ่นมากขึ้น ในการศึกษานี้จะใช้ข้อมูลสภาพอากาศจาก สถานีอุตุนิยมวิทยาสองสถานีได้แก่ สถานีเชียงใหม่ และสถานีอุบลราชธานี เป็นข้อมูลตั้งแต่ปี 2006 ถึง 2011 ซึ่งจะมีการเก็บข้อมูลสถิติสภาพอากาศเป็นคาบ ทาบละ 1 ชั่วโมง

1.2 วัตถุประสงค์

- 1.2.1 พัฒนารูปแบบการคำนวณหาความต้องการน้ำของพืชด้วยวิธี Machine Learning
- 1.2.2 พัฒนารูปแบบการคำนวณหาความต้องการน้ำของพืชที่มีค่าสูญหายของข้อมูล
- 1.2.3 เปรียบเทียบประสิทธิภาพในการหาความต้องการน้ำของพืชในแบบต่างๆ

1.3 ขอบเขตงานวิจัย

- 1.3.1 ศึกษาการคำนวณหาความต้องการน้ำของพืช
- 1.3.2 ศึกษากระบวนการเรียนรู้ของคอมพิวเตอร์ด้วยวิธี Machine Learning
- 1.3.3 เปรียบเทียบข้อมูล จากวิธีการหาความต้องการน้ำของพืชด้วยวิธี Machine Learning กับวิธีทั่วไป

1.4 วิธีการดำเนินงาน

- 1.4.1 จัดหาข้อมูลสภาพแวดล้อมทางภูมิอากาศ
- 1.4.2 นำข้อมูลสภาพแวดล้อมทางภูมิอากาศมาคำนวณหาความต้องการน้ำของพืชด้วยวิธีทั่วไป
- 1.4.3 นำข้อมูลสภาพแวดล้อมทางภูมิอากาศมาคำนวณ โดยใช้วิธี Machine Learning
- 1.4.4 นำข้อมูลสภาพแวดล้อมทางภูมิอากาศมาคำนวณ โดยมีการพิจารณาค่าสูญหายด้วย
- 1.4.5 สรุปและวิจารณ์ผลการทดลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

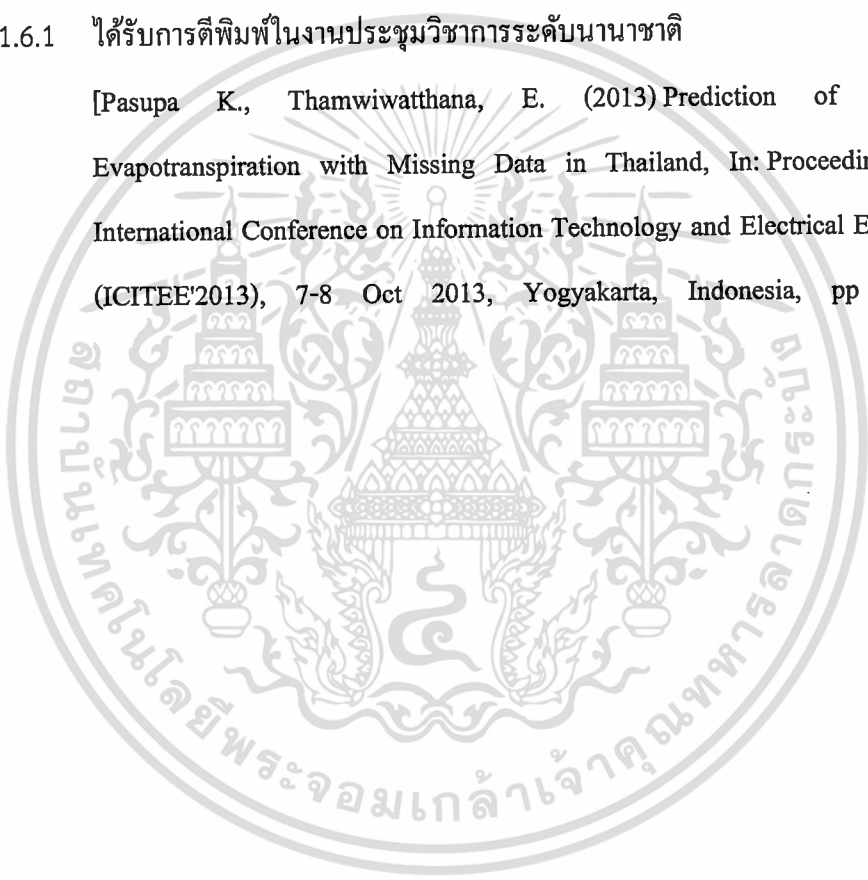
1.5 ประโยชน์ที่คาดว่าจะได้รับ

- 1.5.1 สามารถหาความต้องการน้ำของพืชด้วยวิธี Machine Learning ได้
- 1.5.2 แก้ปัญหาการหาความต้องการน้ำของพืชจากข้อมูลสภาพอากาศที่มีค่าสูญหาย
- 1.5.3 สามารถเปรียบเทียบการหาความต้องการน้ำของพืชด้วยวิธี Machine Learning และวิธีทั่วไป

1.6 ผลลัพธ์จากการศึกษา

- 1.6.1 ได้รับการตีพิมพ์ในงานประชุมวิชาการระดับนานาชาติ

[Pasupa K., Thamwivatthana, E. (2013) Prediction of Reference Evapotranspiration with Missing Data in Thailand, In: Proceeding of 5th International Conference on Information Technology and Electrical Engineering (ICITEE'2013), 7-8 Oct 2013, Yogyakarta, Indonesia, pp 181-186.



บทที่ 2

บททวนวรรณกรรมและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีและหลักการต่างๆ รวมถึงเทคโนโลยีที่เกี่ยวข้องในการพัฒนาระบบ ซึ่งเนื้อหาเกี่ยวข้องกับความต้องการน้ำของพืชและงานวิจัยที่เกี่ยวข้อง และเทคโนโลยีปัญญาประดิษฐ์ ดังต่อไปนี้

2.1 งานวิจัยที่เกี่ยวข้อง

ประเทศไทยนั้นมีการริเริ่มในการใช้เทคโนโลยีสารสนเทศ โดยมีศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) ให้การสนับสนุนซึ่งมีโครงการ Smart Farm Flagship เพื่อใช้เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์สารสนเทศเข้ามาประยุกต์ใช้ในการเกษตร เพื่อยกระดับผลิตภัณ์ มาตรฐานสินค้า และลดต้นทุน โดยพัฒนาเกษตรกรรมใน 4 ด้าน [5] คือ

1. ลดต้นทุน
2. เพิ่มคุณภาพการผลิตและมาตรฐานสินค้า
3. ลดความเสี่ยงจากศัตรูพืชและภัยธรรมชาติ
4. การจัดการและส่งผ่านความรู้

ซึ่งได้มีการนำเทคโนโลยีที่เกี่ยวข้องกับ Smart Farm มาใช้งาน เช่น การเกษตรความแม่นยำสูง (Precision Agriculture) มาใช้ในแปลงพืชซึ่งมีการประยุกต์ใช้งานจริง เช่น ในไร่ร่องนกรานมอนเต้ จังหวัดนครราชสีมา มีระบบ ไร่ร่องน ไวนอัจฉริยะ ใช้ระบบตรวจสอบสภาพอากาศในพื้นที่ขนาดเล็ก (Micro-climate) มาใช้ภายในไร่ เพื่อค้นหาศัตรูพืชและ เพื่อหาความต้องการ การใช้ทรัพยากรของพืชได้อย่างแม่นยำ เพื่อช่วยการจัดการทรัพยากรที่เหมาะสม การหาความต้องการน้ำของพืชนั้นมีความสำคัญเป็นอย่างยิ่งในการทำเกษตรยุคใหม่เพื่อแก้ไขปัญหาการขาดแคลนน้ำ ได้มีงานวิจัยที่เกี่ยวข้องที่พยายามหาความต้องการน้ำของพืชได้อย่างแม่นยำด้วยการใช้ปัญญาประดิษฐ์เพื่อการคำนวณความต้องการน้ำทางการเกษตร โดยวิธีการที่แตกต่างกันไป เพื่อสร้างรูปแบบของการพยากรณ์ที่มีความแม่นยำ ยกตัวอย่างเช่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในปี 2008 งานของ S. K. Jain และคณะ [6] ได้นำเสนอโมเดลที่ใช้ในการพยากรณ์หาความต้องการปริมาณการใช้น้ำของพืชอ้างอิง โดยอาศัยข้อมูลทางภูมิอากาศและโครงข่ายระบบประสาทเทียมเพื่อหาข้อเปรียบเทียบระหว่าง โมเดลแต่ละโมเดลที่ทำการทดลอง ในการทดลองได้มีการจำกัดชนิดของข้อมูลของสภาพอากาศ โดยพบว่า ในการหาความต้องการน้ำอ้างอิงของพืช เพื่อหาความต้องการน้ำของพืช ตัวแปรที่มีความสำคัญและส่งผลต่อความแม่นยำที่สุด ในการใช้ในการคำนวณหาความต้องการน้ำอ้างอิงของพืช คืออุณหภูมิและการแผ่รังสีของดวงอาทิตย์ มีผลต่อการปรับน้ำหนักของนิวรอน (Neuron) โดยพบว่า โมเดลที่ใช้อุณหภูมิ ความชื้น กับ โมเดลที่ใช้อุณหภูมิ ความชื้นและความเร็วลม นั้นไม่สามารถให้ผลการประเมินความต้องการน้ำของพืชอ้างอิงที่ดีที่สุดได้ และโมเดลที่สามารถใช้ในการประเมินความต้องการน้ำของพืชอ้างอิงที่ดีที่สุดโดยใช้โครงข่ายประสาทเทียมคือ อุณหภูมิ ความชื้น จุดน้ำค้าง การแผ่รังสีของดวงอาทิตย์ และความเร็วลม

ในปี 2008 จากงานของ Y. M. Wang และคณะ [7] ได้ศึกษาเกี่ยวกับการหาความต้องการน้ำของพืชด้วยการใช้โครงข่ายประสาทเทียม ในประเทศบรูไนพาโซ โดยใช้เฉพาะข้อมูลอุณหภูมิเป็นข้อมูลเข้าสู่ระบบ เพื่อใช้แก้ปัญหาในประเทศกำลังพัฒนาที่ขาดข้อมูลสภาพอากาศเพื่อใช้ในการจัดการทางการเกษตร ซึ่งสามารถสรุปได้ว่าวิธีการแบบ Penman-Monteith (PM) โดยใช้โครงข่ายประสาทเทียมมาช่วยในการประเมินค่าความต้องการน้ำของพืชในสถานะที่มีข้อมูลสภาพอากาศอย่างจำกัดนั้นมีประสิทธิภาพ โดยแนวทางนี้ช่วยให้สามารถช่วยในการจัดการทางชลประทานในโครงการขนาดเล็กที่มีข้อมูลอย่างจำกัดเมื่อขาดข้อมูลที่เกี่ยวข้องกับความชื้นสัมพัทธ์ ความเร็วลม และข้อมูลของแสงอาทิตย์ได้

งานของ C. Archana และ Shrivastava R.K. [8] ในปี 2010 ได้นำเสนอการใช้ Feed-forward Neural Network (FFNN) เพื่อประเมินหาความต้องการของน้ำแบบรายเดือน โดยนำค่าในช่วงเดือนมกราคม 1986 ถึง เดือนธันวาคมปี 2005 โดยอ้างอิงการประเมินความต้องการด้วยวิธีแบบ PM โดยเปรียบเทียบการคำนวณแบบการถดถอยพหุคูณ กับวิธีการใช้โครงข่ายประสาทเทียมแบบต่างๆ พบว่า ระบบโครงข่ายประสาทเทียม นั้นเหมาะสมที่ใช้ในการหาค่าความต้องการน้ำของพืชอ้างอิง โดยการใช้ข้อมูลสภาพอากาศแบบจำกัดได้ผลที่ดีกว่าการใช้วิธีการแบบการถดถอยพหุคูณ

2.2 ความต้องการน้ำของพืช

ความต้องการน้ำของพืช มีปัจจัยหลายอย่างที่เกี่ยวข้อง ได้แก่ ดิน น้ำ ความชื้นของดิน จุดเหี่ยวเฉาถาวร ความชื้นที่พืชนำไปใช้ได้ การเคลื่อนที่ของน้ำในดิน แรงดึงความชื้น แต่สิ่งที่สมควรเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทราบอีกก็คือ ระยะเวลาการให้น้ำ การคายระเหย การหาปริมาณการใช้น้ำของพืช ความต้องการน้ำของพืช เป็นต้น

2.2.1 ปริมาณการใช้น้ำของพืช

ปริมาณการใช้น้ำของพืช คือปริมาณน้ำที่เกิดจากการคายน้ำของพืช และปริมาณการระเหยของน้ำจากดิน ในการหาความต้องการการใช้น้ำของพืช จะมีอัตราส่วนของการคายน้ำกับการระเหย ซึ่งจะหาได้จากสมการ

$$ET_c = ET_0 \cdot K_c \quad (2.1)$$

เมื่อ

ET_c = อัตราการใช้น้ำของพืช (ลิตร/ต้น/วัน)

ET_0 = อัตราการคายน้ำอ้างอิงของพืช (ลิตร/ต้น/วัน) ซึ่งสามารถหาได้จากข้อมูลของสภาพอากาศ

K_c = เป็นค่าสัมประสิทธิ์การใช้น้ำของพืชแต่ละชนิด

2.2.2 การระเหยและการคายระเหย

การระเหย (Evaporation) เป็นกระบวนการที่เกิดขึ้นจากน้ำเปลี่ยนสภาพจากของเหลวกลายเป็นไอและหายไปจากพื้นผิว ซึ่งมีการใช้พลังงานในการเปลี่ยนสถานะของโมเลกุลน้ำในการเปลี่ยนจากของเหลวไปเป็นไอ เช่น การแผ่รังสีของดวงอาทิตย์หรืออุณหภูมิของสภาพอากาศ เป็นต้น

การคายระเหย (Evapotranspiration) เป็นผลรวมของการใช้น้ำของพืช ที่พืชสูญเสียสู่อากาศ ในรูปของไอน้ำ ประกอบไปด้วยส่วนหลักสองส่วนคือ น้ำจากการคายน้ำของพืช และการระเหยของน้ำจากผิวดิน ซึ่งขึ้นอยู่กับชนิดของพืชที่ทำการปลูกและชนิดของดิน ซึ่งองค์ประกอบสำคัญที่มีผลต่อการคายระเหยและระดับการเจริญเติบโตของพืชคือ ร้อยละของหน้าดินที่ปกคลุม การแผ่รังสีของดวงอาทิตย์ ความชื้นของดินอุณหภูมิ และกระแสลม

2.2.3 ปัจจัยของสภาพอากาศ

ปัจจัยของสภาพอากาศที่ส่งผลกระทบต่อการคายระเหยของพืชนั้นประกอบไปด้วย การแผ่รังสีของดวงอาทิตย์ อากาศ อุณหภูมิ ความชื้นของดิน และความเร็วลม กระบวนการที่นำมาใช้เพื่อหาค่าอัตราการระเหยน้ำ ซึ่งปริมาณการคายระเหยของพืช นั้นจะใช้สภาพของสิ่งแวดล้อมมาคำนวณ

2.2.4 ค่าสัมประสิทธิ์พืช

ค่าสัมประสิทธิ์พืช (K_c) เป็นค่าคงที่ของพืชที่ได้จากความสัมพันธ์ของปริมาณการใช้น้ำที่ทำการทดลองและตรวจวัดจากถาดวัดการใช้น้ำของพืช กับผลการคำนวณปริมาณการคายระเหยของพืชอ้างอิง ซึ่งมีความจำเป็นในการนำไปใช้ด้านการเกษตร นำมาคำนวณกับข้อมูลสภาพอากาศเพื่อหาค่าที่ถูกต้อง และนำไปใช้งานได้อย่างมีประสิทธิภาพ ค่าสัมประสิทธิ์พืชจะมีค่าที่แตกต่างซึ่งขึ้นอยู่กับชนิดและช่วงเวลาในการเติบโตของพืช

2.3 โมเดลที่ใช้ในการคำนวณหาค่าการคายน้ำอ้างอิงของพืช

มากกว่า 50 ปีที่มีการพยายามพัฒนากระบวนการหาการประมาณค่าการคายน้ำอ้างอิงของพืช มีงานวิจัยและกระบวนการที่เกี่ยวข้องจำนวนมาก เช่น วิธีของ Blaney-Criddle, Radiation, Penman, Modified Penman และ Pan Evaporation ซึ่งวิธี Modified Penman นั้นถูกเสนอว่าเป็นวิธีการที่ดีที่สุดในการหาค่าการคายน้ำอ้างอิงของพืช มีอัตราความผิดพลาดน้อยที่สุด ใช้ข้อมูลการวัดของอุณหภูมิของอากาศและแสงแดดหรือการแผ่รังสี แต่วิธี Modified Penman นั้นจะพบได้ว่ามีความผิดพลาดได้ถึงร้อยละ 20 เมื่อสภาพอากาศนั้นมีความชื้นต่ำ ส่วนวิธีการแบบอื่นนั้น วิธี Blaney-Criddle ถูกแนะนำให้ใช้ในการคำนวณแบบคาบเวลาที่ยาวกว่า 1 เดือน วิธี Pan evaporation นั้นถูกแนะนำให้ใช้กับคาบเวลาที่ยาวกว่า 10 วัน วิธี Radiation นั้นจะมีผลลัพธ์ที่ดีในสภาวะอากาศแบบชื้น องค์การอาหารและการเกษตรแห่งสหประชาชาติ (FAO) นั้นได้วิเคราะห์ประสิทธิภาพวิธีการหาค่าการคายระเหยของพืชแบบต่างๆ และวิธี Penman-Monteith (PM) นั้นได้ถูกแนะนำเป็นวิธีการมาตรฐานเพียงวิธีเดียว เนื่องจากเป็นวิธีการที่ใช้ได้ในพื้นที่ที่กว้างขวางและสามารถใช้ในการคำนวณในช่วงสั้นๆ ได้ โดยใช้การแผ่รังสีของดวงอาทิตย์อุณหภูมิของอากาศ ความชื้นของอากาศ และข้อมูลความเร็วของกระแสลม [5] วิธี PM ดังสมการที่ 2.2

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} U_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \quad (2.2)$$

ความหมาย

- ET_0 = ปริมาณการใช้น้ำของพืชอ้างอิง (มม./วัน)
- R_n = ปริมาณรังสีของดวงอาทิตย์ทั้งหมดที่พืชได้รับ ($MJ/m^2/d$)
- G = Flux ค่าความร้อนของพื้นดิน ($MJ/m^2/d$)
- T = อุณหภูมิของอากาศเฉลี่ย ($^{\circ}C$)
- Δ = ค่าความลาดเทของเส้น curve แรงดันไอ ($kPa/^{\circ}C$)
- γ = ค่าคงที่ของ Psychometric ($kPa/^{\circ}C$)
- U_2 = ค่าความเร็วลมที่ระดับความสูงจากพื้นดิน 2 ม. (m/s)
- $(e_s - e_a)$ = ค่าความต่างของแรงดันไอ (kPa)

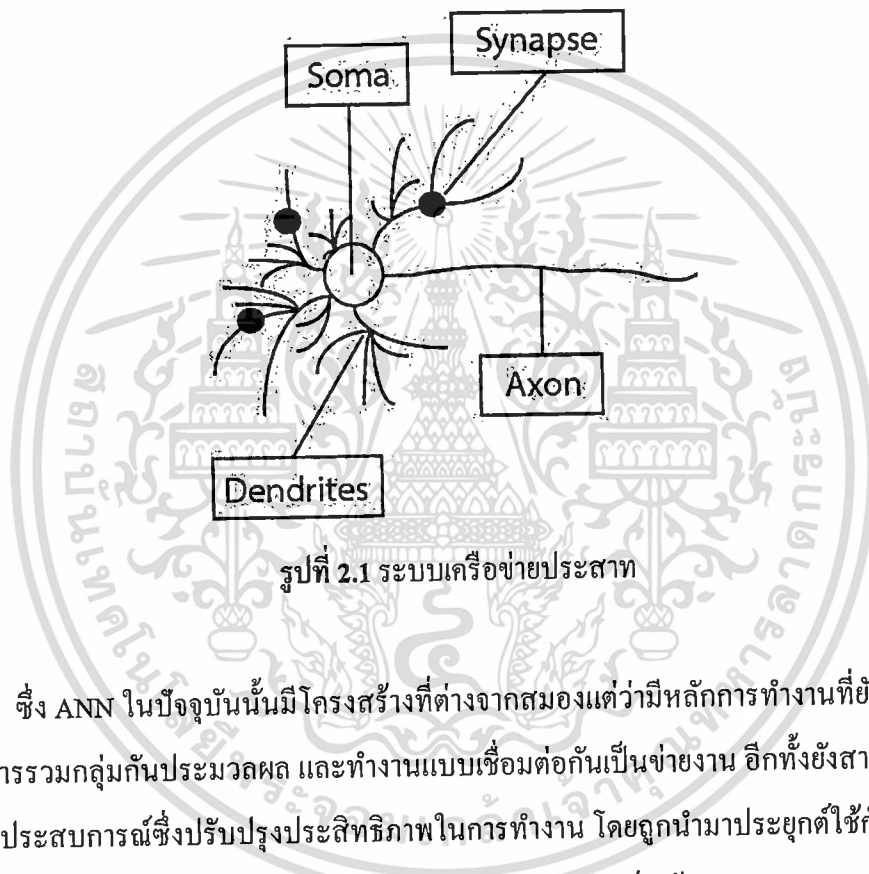
2.4 การเรียนรู้ของเครื่อง (Machine Learning)

เป็นวิธีการหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligent) ซึ่งเกี่ยวกับการพัฒนาวิธีการเพื่อให้คอมพิวเตอร์สามารถเรียนรู้ได้จากข้อมูล เพื่อเพิ่มประสิทธิภาพในการแก้ปัญหาต่างๆ เช่น การจำแนกข้อมูลของ ภาพ และ เสียง แยกแยะหรือแบ่งกลุ่มข้อมูลจำนวนมาก ด้วยอัลกอริทึมจากการใช้ข้อมูลชุดสอน (Training Data) เพื่อให้เครื่องสามารถสร้างรูปแบบการแยกแยะหรือแบ่งกลุ่มข้อมูลได้ โดยมีขั้นตอนวิธีการดังต่อไปนี้

1. การเรียนรู้แบบมีผู้สอน (Supervised Learning) ใช้การเรียนรู้ข้อมูลจากข้อมูลที่มีอยู่แล้ว เพื่อให้คอมพิวเตอร์เข้าใจรูปแบบของข้อมูล เพื่อจัดการกับข้อมูลที่จะเข้ามาในอนาคต
2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เรียนรู้ด้วยการนำข้อมูลไปจัดกลุ่ม ด้วยความคล้ายคลึงกันของข้อมูล จนได้กลุ่มที่เหมาะสม เรียกว่าการแบ่งกลุ่ม (Cluster)
3. การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) เป็นการเรียนรู้จากการสังเกตได้จากสภาพแวดล้อม หรือลองผิดลองถูก

2.5 ระบบโครงข่ายประสาทเทียม (Artificial Neural Network)

ระบบโครงข่ายประสาทเทียม (ANN) หรือที่นิยมเรียกว่า Neural Networks เป็นวิธีการเรียนรู้แบบมีผู้สอน ถูกนำมาจากการทำงานที่อยู่ภายในสมองมนุษย์ที่ใช้ในการคำนวณซึ่งมีความซับซ้อน ซึ่งจำลองการทำงานของเซลล์สมอง โครงสร้างของ Neural นั้นเป็นโครงสร้างพื้นฐานแบบง่าย ๆ ซึ่งเมื่อนำมารวมกันแล้วสามารถใช้คำนวณได้อย่างมีประสิทธิภาพซึ่ง จะประกอบไปด้วยส่วนที่เรียกว่า Soma ซึ่งเป็นตัวเซลล์ มีส่วนของใยประสาทที่ยื่นออกไปเรียกว่า Dendrite และมีสายที่เชื่อมต่อกับเซลล์อื่นเรียกว่า Axon ดังรูปที่ 2.1

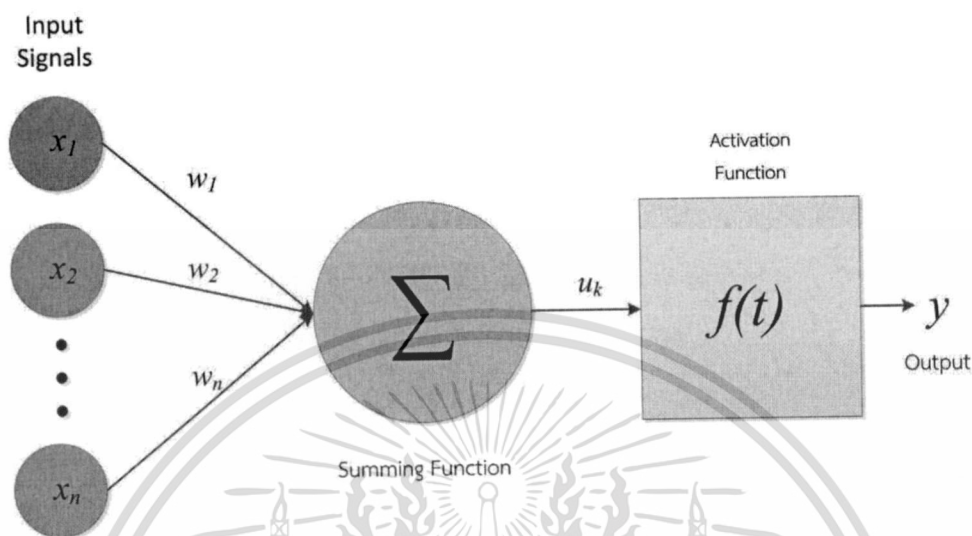


รูปที่ 2.1 ระบบเครือข่ายประสาท

ซึ่ง ANN ในปัจจุบันนั้นมีโครงสร้างที่ต่างจากสมองแต่ยังมีหลักการทำงานที่ยังเหมือนกัน คือ มีการรวมกลุ่มกันประมวลผล และทำงานแบบเชื่อมต่อกันเป็นข่ายงาน อีกทั้งยังสามารถเรียนรู้ได้จากประสบการณ์ซึ่งปรับปรุงประสิทธิภาพในการทำงาน โดยถูกนำมาประยุกต์ใช้กับงานต่างๆ เช่น การจดจำลายมือ การรู้จำเสียง การคำนวณทางอุทุนิยมวิทยา เป็นต้น

ANN นั้นจะมี Node ที่เรียกว่า นิวรอน (Neuron) ซึ่งทำหน้าที่เหมือนกับนิวรอนในทางสมอง ซึ่งทำหน้าที่รับค่ารับเข้า (Input) ที่ถูกนำมาคำนวณกับค่าถ่วงน้ำหนัก (Weight) แล้วถูกส่งเข้ามาสู่นิวรอน และรับข้อมูลเข้ามาโดยค่าป้อนเข้ามาสู่นิวรอน อาจจะตั้งแต่ 1 หรือมากกว่า นำมาหาผลรวมของค่ารับเข้าทั้งหมด ต่อมาส่งค่าผลลัพธ์ออกไปให้กับ Node ที่เชื่อมต่ออยู่ด้วยกัน ANN มีการเชื่อมต่อโดยจะมีการถ่วงน้ำหนักด้วย ผลของค่าผลลัพธ์ที่ออกจากนิวรอน คือผลรวมของการ

กฎ ระหว่างค่ารับเข้าและค่าถ่วงน้ำหนักของแต่ละนิวรอน โดยมีค่าขีดแบ่ง (Threshold) ควบคุมค่าของผลลัพธ์ ไว้ดังรูปที่ 2.2



รูปที่ 2.2 รูปแบบการคำนวณของ Neural Network

ซึ่งในทางคณิตศาสตร์แล้ว เราจะสามารถอธิบาย ภาพโครงสร้างของ นิวรอนออกมาเป็นสมการที่ 2.3 และ 2.4

$$u_k = \sum_{j=1}^m w_j x_j \quad (2.3)$$

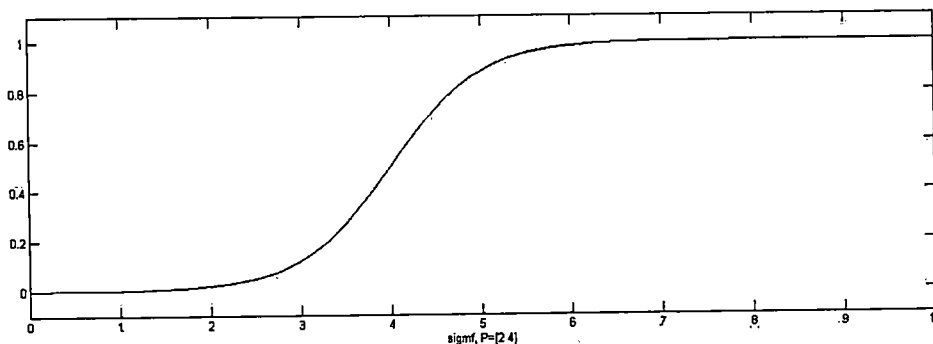
เมื่อ x_1, x_2, \dots, x_m เป็น ค่าป้อนเข้า w_1, w_2, \dots, w_m เป็นค่าน้ำหนักของนิวรอนแล้ว u_k มีค่าเท่ากับผลลัพธ์ของนิวรอนและ y เป็น ค่าของผลลัพธ์ของนิวรอนที่ผ่านการใช้ Activation Function นั้นมีผลทำให้ค่าของผลลัพธ์ เปลี่ยนไปจาก u_k ซึ่งในที่นี้ใช้ Sigmoid Function ซึ่งเป็นฟังก์ชันทางคณิตศาสตร์อย่างหนึ่ง เมื่อนำผลลัพธ์มาวาดบนกราฟจะมีลักษณะคล้ายตัว S ตามสมการที่ 2.5 ดังรูปที่ 2.3

$$y = f(u_k) \quad (2.4)$$

โดยที่ Sigmoid Function มีสมการดังสมการที่ 2.5

$$f(t) = \frac{1}{1 + e^{-t}} \quad (2.5)$$

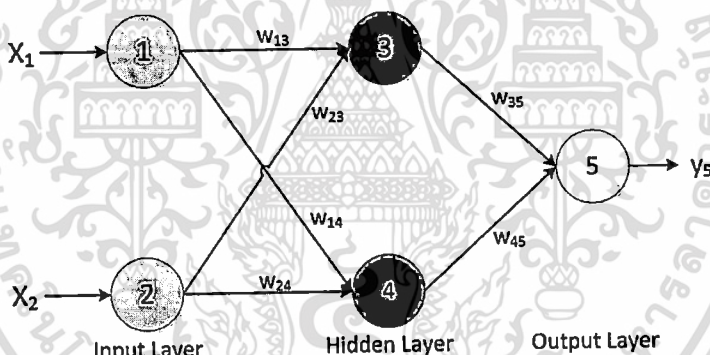
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.3 แสดงลักษณะของ Sigmoid Function

2.6 Back-Propagation Algorithm

การแพร่กระจายแบบย้อนกลับนั้น (Back-Propagation) มีไว้สำหรับการใช้โครงข่ายประสาทเทียมแบบไปข้างหน้าหลายชั้น (Multilayer Feedforward Network) โดยมีจุดประสงค์เพื่อหาค่าต่ำสุดของความผิดพลาดของค่าผลลัพธ์ มีลักษณะโครงสร้างตามรูปที่ 2.4



รูปที่ 2.4 แสดงโครงสร้างของ โครงข่ายประสาทเทียม

การฝึกโครงข่ายประสาทเทียมเพื่อให้เกิดการเรียนรู้จากประสบการณ์มีขั้นตอนดังต่อไปนี้ เริ่มจากการสุ่มค่านำหนักให้กับค่าป้อนเข้าของแต่ละนิวรอน ซึ่งค่านี้สามารถเปลี่ยนแปลงได้ หลังจากทำการฝึกโครงข่ายประสาทเทียมแล้ว การกำหนดค่า Bias ในตอนเริ่มต้น โดยการสุ่มขึ้นมาซึ่งหลังจากทำการฝึกแล้วจะสามารถเปลี่ยนแปลงค่า Bias ได้ ในการคำนวณในแต่ละชั้น ค่าผิดพลาดสามารถนำมาคำนวณได้จากการเทียบผลลัพธ์ที่คำนวณได้กับค่าเป้าหมาย (Target value) ที่ตั้งไว้แล้วทำการปรับน้ำหนักและค่า Bias จากค่าความผิดพลาดที่หาได้ จึงทำซ้ำกระบวนการทั้งหมดเพื่อให้เกิดความถูกต้องของข้อมูลจนค่าความผิดพลาดนั้นหมดไปหรือน้อยจนยอมรับได้

หรือครบจำนวนรอบที่กำหนดโดยที่การกระจายแบบย้อนกลับจะทำการปรับค่าน้ำหนักของแต่ละนิวรอน ซึ่งจะแทนค่าย้อนกลับด้วยเครื่องหมาย δ โดยมีสมการดังสมการที่ 2.6

$$\delta = z - y \quad (2.6)$$

โดย

- δ = ค่าย้อนกลับ
- z = ค่าเป้าหมาย (Target Value)
- y = ผลลัพธ์ของ Output Layer

ซึ่งจะมาทำการคำนวณเพื่อปรับน้ำหนักของแต่ละนิวรอนต่อไปด้วยสมการที่ 2.7

$$\Delta w_{jk} = \alpha \cdot y_j \cdot \delta_k \quad (2.7)$$

โดย

- Δw_{jk} = ค่าน้ำหนักของนิวรอน
- α = อัตราการเรียนรู้ (Learning Rate)
- y_j = ค่าผลลัพธ์ของนิวรอน
- δ_k = ค่าย้อนกลับ

2.7 ความถดถอยเชิงเส้น (Linear Regression)

ความถดถอยเชิงเส้น (LR) ใช้เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรระหว่างตัวแปร y กับตัวแปรอิสระ x เพียงตัวเดียว สนใจศึกษาปัจจัยที่มีอิทธิพลหรือผลต่อตัวแปรตาม y เพียงปัจจัยเดียว แต่โดยทั่วไป ปัจจัยที่มีอิทธิพลต่อ y จะมีหลายปัจจัยหรือกล่าวได้ว่ามีตัวแปรอิสระหลายตัวที่มีอิทธิพลต่อ y [10] ดังสมการที่ 2.8

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (2.8)$$

โดย

- y_i = ค่าของตัวแปรตาม
- β_0 = ส่วนตัดแกน Y
- $\beta_1, \beta_2, \dots, \beta_k$ = สัมประสิทธิ์ความถดถอยเชิงเส้น
- x_k = ตัวแปรอิสระ
- ε = ค่าความคลาดเคลื่อน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.8 การแก้ปัญหาค่าสูญหายของข้อมูล

ปัญหาการสูญหายของข้อมูล เป็นปัญหาที่มักจะพบได้บ่อยในการเก็บข้อมูลทางสถิติในหลายสาขาเพื่อนำมาการทำวิจัย ซึ่งการสูญหายของข้อมูล มีวิธีการทางสถิติซึ่งสามารถแก้ปัญหาค่าสูญหายของข้อมูลอยู่หลายวิธี แต่โดยปกติแล้วมีกระบวนการหลักอยู่ 3 วิธีด้วยกันคือ

1. ลบชุดข้อมูลที่มีค่าสูญหายทิ้งไป ซึ่งมีข้อเสียคือจะทำให้ชุดข้อมูลสูญหายไปเลย
2. Parameter estimation ใช้ในการหา Maximum Likelihood เพื่อประมาณค่ามาแทนค่าสูญหาย
3. Imputation Techniques เช่นการใช้ ค่าเฉลี่ย หรือ ค่ามัธยฐาน เพื่อหาค่ามาทดแทนค่าสูญหาย

2.8 Expectation Maximization Algorithm

Expectation-Maximization Algorithm (EM) เป็นวิธีการที่ใช้ในการแก้ปัญหาค่าสูญหาย โดยการใช้โมเดลของข้อมูล โดยที่ไม่เสียชุดข้อมูลไปเหมือนกับในกรณีของการลบชุดข้อมูลทิ้ง

EM ใช้ค่าค้นหาค่าตัวแปรที่มีค่า Likelihood สูงสุดเพื่อใช้แก้ปัญหาทางสถิติ เมื่อเกิดปัญหาไม่ทราบค่าข้อมูลของข้อมูลหรือมีค่าของข้อมูลหายไป การหาค่าสูงสุดนั้นต้องใช้ ฟังก์ชัน Likelihood กับค่าที่ไม่ทราบค่าทั้งหมด โดยตั้งค่าตัวพารามิเตอร์ให้เหมาะสม หาค่าความน่าจะเป็นที่ต้องการเพื่อให้ผลลัพธ์ใกล้เคียงกับค่าที่ควรจะเป็นมากที่สุด ดังนั้น Likelihood คือตัววัดค่าพารามิเตอร์ว่า พารามิเตอร์ชุดใดมีความเหมาะสมมากกว่ากัน โดยใช้การสร้างแบบจำลองขึ้นมาแล้วนำมาคำนวณซ้ำๆ เพื่อให้ได้ค่าที่ใกล้เคียงกับค่าที่สูญหายไป

EM มีหลักการทำงานอยู่สองขั้นตอนคือ E-step และ M-step โดยเริ่มต้นจากการประมาณค่าตัวแปรความแปรปรวน และค่าเฉลี่ย เพื่อนำมาสร้างสมการความถดถอยสำหรับตัวแปรที่มีค่าสูญหาย ในขั้นตอนต่อไปคือ M-step โดยการใช้สมการเพื่อทดแทนค่าสูญหาย แล้ววนกลับไปใช้ E-step ด้วยค่าเฉลี่ยและความแปรปรวนใหม่ แล้วนำมาสร้างสมการความถดถอยเพื่อหาค่าทดแทนค่าสูญหายอีกครั้ง แล้ววนกลับไปใช้ M-step อีกครั้ง จนกว่าค่าผลลัพธ์จะเกิดการลู่เข้า ซึ่งสมการ E-step มีดังต่อไปนี้

$$Q(\theta | \theta^{(n)}) = E_{z|x, \theta^{(n)}} [\log L(\theta; x, z)] \quad (2.9)$$

และขั้นตอน M-step มีสมการดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\theta^{(t+1)} = \operatorname{argmax} Q(\theta | \theta^{(t)}) \quad (2.10)$$

โดยที่

θ = เวกเตอร์ของพารามิเตอร์

x = ข้อมูลค่าสังเกต

z = แทนค่าสูญเสีย

เริ่มจากสุ่มค่าตัวแปรให้ θ แล้วนำมาคำนวณหาค่า z โดยให้ค่าพารามิเตอร์ เพื่อหาค่าที่ดีที่สุดใช้สำหรับพารามิเตอร์ θ ซึ่งมีความสัมพันธ์กับค่า z แล้วทำซ้ำไปเรื่อยๆ ด้วยการคำนวณหาค่า z อีกครั้งจนกว่าค่าจะเกิดการลู่เข้า

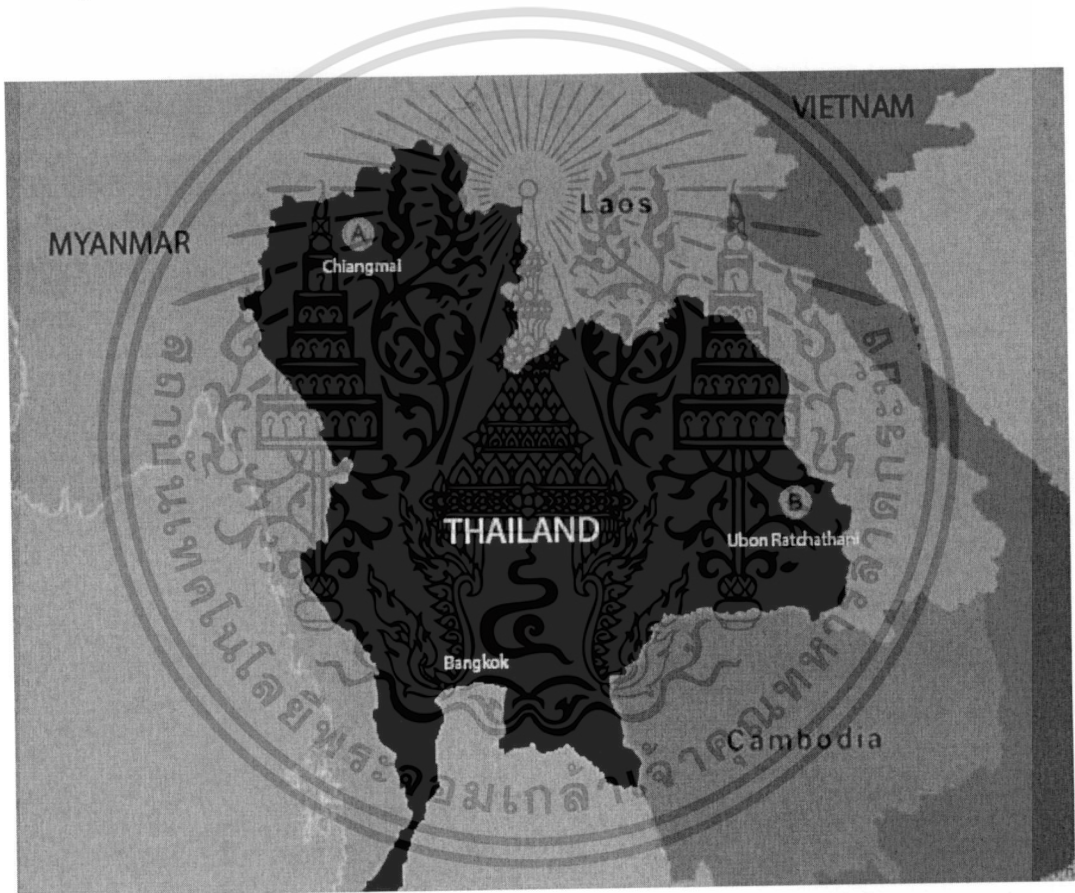


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

การออกแบบการทดลอง

ในงานวิจัยนี้ได้มีการออกแบบการทดลองเพื่อหา ความต้องการน้ำอ้างอิงของพืช โดยการใช้ข้อมูลที่ถูกเก็บจากสถานีตรวจอากาศจังหวัดเชียงใหม่ และจังหวัดอุบลราชธานี ซึ่งจังหวัดเชียงใหม่ตั้งอยู่ในพื้นที่ภาคเหนือ และจังหวัดอุบลราชธานีภาคตะวันออกเฉียงเหนือของประเทศไทย ดังรูปที่ 3.1



รูปที่ 3.1 แสดงแผนที่จังหวัดเชียงใหม่ และจังหวัดอุบลราชธานี

ซึ่งข้อมูลจากทั้งสองสถานีนั้นมีการเก็บข้อมูลสภาพอากาศตั้งแต่ปี พ.ศ. 2549 ถึง พ.ศ. 2554 มีการจัดเก็บรายชั่วโมง ในชุดข้อมูลสภาพอากาศมีค่าสูญหายอยู่จำนวน 2.2% ของข้อมูลทั้งหมด ทำให้ไม่สามารถทำการคำนวณหาค่าความต้องการน้ำอ้างอิงของพืชได้เป็นจำนวน 11.8% ของชุดข้อมูลทั้งหมด ซึ่งมีจำนวนตัวแปรลักษณะจำเพาะดังตารางที่ 3.1

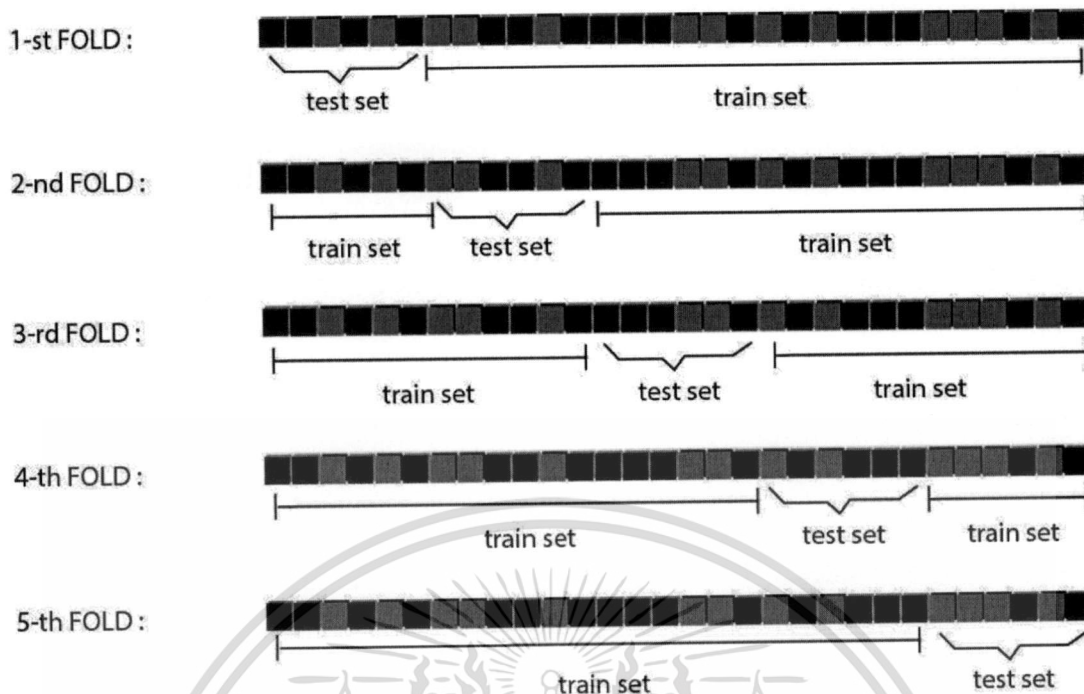
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 แสดงตัวแปรลักษณะจำเพาะของข้อมูลสภาพอากาศที่ใช้ในการหาความต้องการน้ำอ้างอิงของพืช

Index	Feature
1	วันที่ของปี
2	ความสูงจากระดับน้ำทะเล (m)
3	ละติจูด (เรเดียน)
4	ความเร็วลม (m/s ที่ความสูง 2 เมตรจากพื้นดิน)
5	ความยาวนานแสงแดด (ชั่วโมง/วัน)
6	อุณหภูมิสูงสุด (°C)
7	อุณหภูมิต่ำสุด (°C)
8	อุณหภูมิเฉลี่ย (°C)
9	ความชื้นสัมพัทธ์ (%)

ในการวิจัย แบ่งกระบวนการออกเป็น 3 การทดลองดังต่อไปนี้

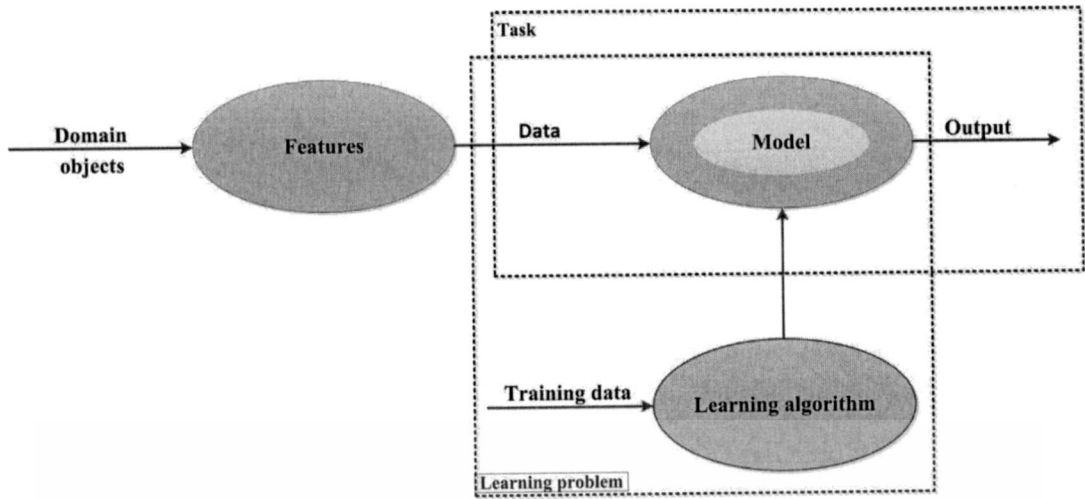
1. การทดลองหาความต้องการน้ำอ้างอิงของพืชโดยอัลกอริทึมต่างๆ โดยแบ่งข้อมูลออกเป็น 5 ชุดในขนาดที่เท่ากัน เพื่อทำการทดสอบด้วยวิธีการ k -fold Cross-Validation ตามตัวอย่างในรูปที่ 3.2 และกำหนดเซตของข้อมูลขึ้นมา 2 เซต เซตที่ 1 เป็นข้อมูลชุดสอน เซตที่ 2 เป็นข้อมูลชุดทดสอบ และทดสอบการคำนวณหาความต้องการน้ำอ้างอิงของพืชด้วยวิธี LR ส่วนการใช้วิธี ANN เพื่อหาคำนวณหาความต้องการน้ำของพืชอ้างอิง จะทำ k -fold Cross-Validation อีกครั้งในข้อมูลชุดสอนเพื่อหาจำนวนโหนดของ Hidden Node ที่ดีที่สุด แล้วนำมาทดสอบกับข้อมูลชุดทดสอบทุกชุด แล้วนำมาหาค่าเฉลี่ย R^2 เพื่อเปรียบเทียบผลลัพธ์กับการคำนวณหาความต้องการน้ำอ้างอิงโดยวิธี PM ในการทำการทดลองนี้ทำขึ้นเพื่อทดสอบว่า การหาความต้องการน้ำอ้างอิงของพืชนั้นเป็นแบบไม่เชิงเส้น ANN เป็นวิธีการคำนวณแบบไม่เชิงเส้น ส่วน LR เป็นวิธีการคำนวณแบบเชิงเส้น จึงถูกนำมาเปรียบเทียบกัน โดยใช้ผลลัพธ์จากวิธีการคำนวณแบบ PM มาเปรียบเทียบผลลัพธ์การคำนวณ



รูปที่ 3.2 แสดงภาพการทำงานของ 5-fold Cross-Validation

2. การทดสอบหาความต้องการน้ำอ้างอิงของพืชโดยมีค่าสูญหายของข้อมูลสภาพอากาศแบ่งข้อมูลเช่นเดียวกับการทดลองที่ 1 แล้วทำการสุ่มค่าสูญหายจำนวน 5% 10% และ 15% เข้าไปในข้อมูลชุดทดสอบ แล้วนำมาทดแทนค่าสูญหายด้วยวิธี 1. EM และ 2 ใช้การทดแทนค่าสูญหายด้วยวิธีนำค่าในวันที่ผ่านมาแทนที่ จึงนำข้อมูลที่ถูกทดแทนค่าสูญหายแล้ว ไปใช้เป็นข้อมูลชุดทดสอบเพื่อหาความต้องการน้ำอ้างอิงของพืชเช่นเดียวกับในการทดลองที่ 1 ด้วยวิธี ANN และ LR

3. การทดลองเปรียบเทียบการเลือกใช้ตัวแปรข้อมูลลักษณะจำเพาะของสภาพอากาศ โดยใช้ Linear Regression เพื่อหาค่าน้ำหนักของแต่ละตัวแปรลักษณะจำเพาะ นำตัวแปรลักษณะจำเพาะมาทำการทดลองเช่นเดียวกับการทดลองที่ 1 และการทดลองที่ 2 เพื่อหาความแม่นยำ เพิ่มตัวแปรลักษณะจำเพาะเข้าไปทีละ 1 ตัว โดยเริ่มจากตัวแปรลักษณะจำเพาะ 1 ตัว ในการทดลองหาความต้องการน้ำอ้างอิงของพืชมีกระบวนการดังรูปที่ 3.3



รูปที่ 3.3 แสดงภาพรวมของการทำงานของวิธี Machine Learning [9]

ในบทนี้ได้กล่าวถึงการทดสอบหาความต้องการน้ำอ้างอิงของพืชและการทดแทนค่าสูญหาย เพื่อเปรียบเทียบประสิทธิภาพวิธีคำนวณด้วยอัลกอริทึม เปรียบเทียบประสิทธิภาพการคำนวณ เมื่อมีค่าสูญหายในข้อมูลสภาพอากาศ และเปรียบเทียบลักษณะการเลือกใช้ตัวแปรลักษณะจำเพาะ

บทที่ 4

ผลการทดลอง

จากการออกแบบการทดลองในบทที่ 3 ได้ทำการทดลอง และได้ผลลัพธ์ในการหาความต้องการน้ำอ้างอิงของพืชทั้ง 3 การทดลองคือ

1. การหาความต้องการน้ำของพืชอ้างอิงโดยที่ไม่มีค่าสูญหายของสภาพภูมิอากาศ
2. การหาความต้องการน้ำของพืชอ้างอิงโดยมีค่าสูญหายของสภาพภูมิอากาศ
3. การหาความต้องการน้ำของพืชอ้างอิงโดยใช้ตัวแปรลักษณะจำเพาะ

โดยได้ผลลัพธ์ดังต่อไปนี้

4.1 การหาความต้องการน้ำอ้างอิงของพืชโดยที่ไม่มีค่าสูญหายของสภาพภูมิอากาศ

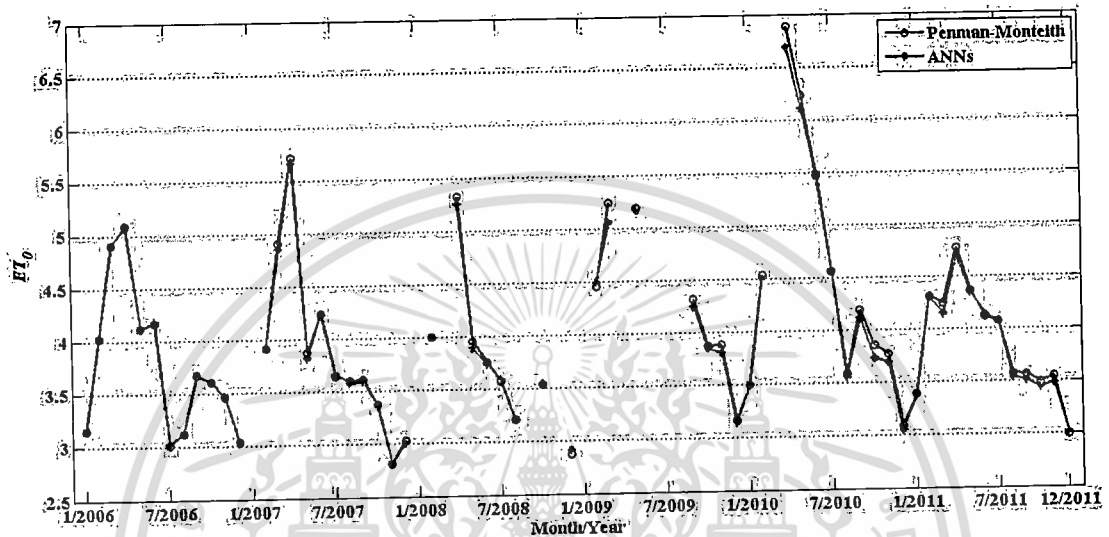
แบ่งข้อมูลสภาพสภาพอากาศด้วยวิธี k -fold Cross-Validation และหาโมเดลที่ดีที่สุดของ ANN ในแต่ละชุดทดสอบ แล้วนำโมเดลมาทดสอบกับชุดทดสอบทุกชุดเพื่อหาโมเดลที่ดีที่สุด ได้ผลลัพธ์ค่าเฉลี่ยความแม่นยำ ดังตารางที่ 4.1

ตารางที่ 4.1 แสดงค่าเฉลี่ยผลลัพธ์การหาความต้องการน้ำอ้างอิงของพืชโดยใช้วิธีการคำนวณแบบ ANN และ LR โดยที่ไม่มีค่าสูญหาย

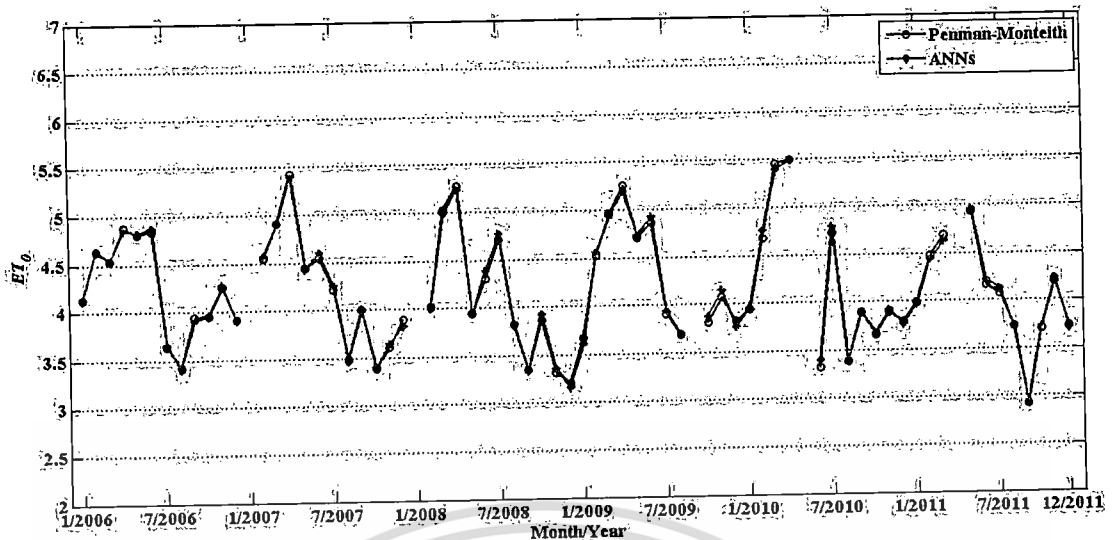
Algorithms	R^2
ANN (9-46-1)	0.9998
LR	0.9372

พบว่าค่าผลลัพธ์ของการหาความต้องการน้ำอ้างอิงของพืช โดยใช้วิธีการคำนวณแบบ ANN ซึ่งมีโครงข่ายแบบ 9-46-1 ซึ่งได้มาจากการทดลองหาจำนวนโหนดใน Hidden Node ที่มีผลลัพธ์ที่ดีที่สุดของข้อมูลชุดสอนนั้นให้ผลลัพธ์ที่ดีกว่าการคำนวณหาความต้องการน้ำอ้างอิงของพืชแบบ LR เพราะว่าการหาความต้องการน้ำอ้างอิงของพืชเป็นโมเดลแบบไม่เชิงเส้น จึงได้ผลลัพธ์ที่ดีกว่า LR ที่เป็นวิธีแบบเชิงเส้น

ขั้นตอนต่อไปทำการทดลองด้วยการเปรียบเทียบการหาความต้องการน้ำอ้างอิงของพืชโดยใช้วิธีการคำนวณแบบ ANN เปรียบเทียบกับการหาความต้องการน้ำอ้างอิงของพืชด้วยวิธี PM จากข้อมูลของสถานีตรวจอากาศเชียงใหม่ และอุบลราชธานีในปี พ.ศ. 2549-2554 (ค.ศ. 2006-2011) ได้รูปที่ 4.1 และ 4.2



รูปที่ 4.1 แสดงการเปรียบเทียบของผลลัพธ์ความต้องการน้ำอ้างอิงของพืชระหว่างวิธีการคำนวณแบบ ANN และวิธี PM จากข้อมูลของสถานีตรวจอากาศจังหวัดเชียงใหม่ในปี พ.ศ. 2549-2554



รูปที่ 4.2 แสดงการเปรียบเทียบของผลลัพธ์ความต้องการน้ำอ้างอิงของพีชระหว่างวิธีการคำนวณแบบ ANN และวิธี PM จากข้อมูลของสถานีตรวจอากาศจังหวัดอุบลราชธานีในปี พ.ศ. 2549-2554

จากรูปภาพที่ 4.1 และ 4.2 จะเห็นได้ว่าค่าความต้องการน้ำอ้างอิงของพีชนั้นมีค่ามากหรือน้อยขึ้นอยู่กับฤดูกาล โดยในช่วงฤดูร้อน (เดือนมีนาคม-เดือนพฤษภาคม) มีค่าความต้องการน้ำอ้างอิงของพีช มักจะสูงที่สุดในปีนั้นๆ และจะลดต่ำลงในช่วงหน้าฝนซึ่งเป็นฤดูกาลถัดไป ซึ่งมีความคล้ายกันทั้งในการหาค่าความต้องการน้ำของพีชอ้างอิงของพีช ที่ใช้ข้อมูลจากสถานีตรวจอากาศจังหวัดเชียงใหม่และอุบลราชธานี

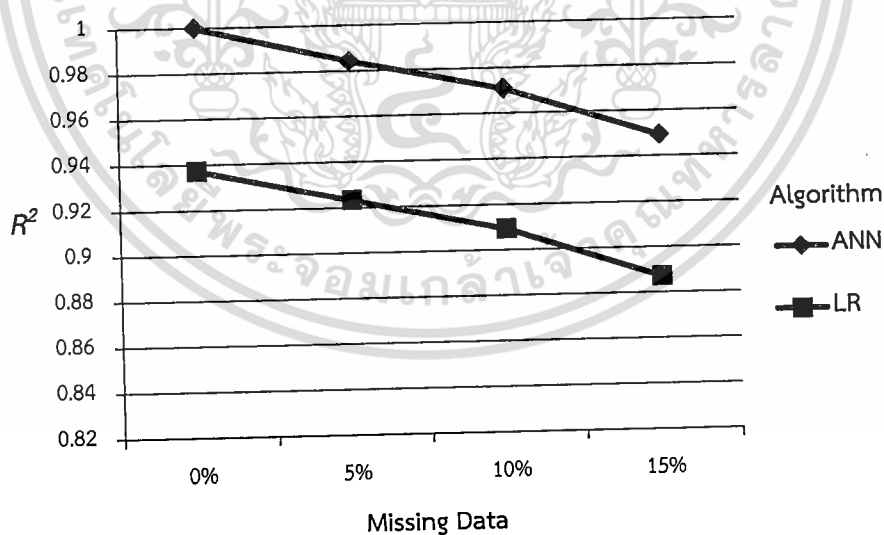
4.2 การหาความต้องการน้ำของพีชอ้างอิงโดยมีค่าสูญหายของสภาพภูมิอากาศ

ทำการแบ่งข้อมูลออกเป็น 5 ชุดเช่นเดียวกับในการทดลองที่ 4.1 และทำการสุ่มค่าสูญหายในข้อมูลชุดทดสอบ ลงไปในข้อมูลจำนวน 5% 10% และ 15% ในตัวแปร 5 ตัวได้แก่ 1. ความเร็วลม 2. ความยาวนานแสงแดด 3. อุณหภูมิต่ำสุด 4. อุณหภูมิสูงสุด 5. ความชื้นสัมพัทธ์ แล้วนำมาทดแทนค่าสูญหายด้วยการใช้วิธีการ EM และ ใช้การทดแทนค่าสูญหายด้วยวิธีนำค่าในวันที่ผ่านมาแทนที่ แล้วนำมาคำนวณหาความต้องการน้ำอ้างอิงของพีชด้วยวิธีการคำนวณแบบ ANN และ LR เหมือนในขั้นตอนที่ 4.1 การทดลองทดแทนค่าสูญหายด้วยวิธี EM ได้ผลลัพธ์ดังตารางที่ 4.2

ตารางที่ 4.2 แสดงค่าเฉลี่ยผลลัพธ์ของการหาความต้องการน้ำอ้างอิงของพืชโดยมีค่าสูญหาย ที่มีการทดแทนค่าสูญหายด้วยวิธี EM เปรียบเทียบระหว่าง วิธีการคำนวณแบบ ANN และ LR

Algorithm	Percentage of Missing Values	R^2 of Re-estimated Features	R^2 of the Prediction
ANN (9-46-1)	5%	0.9762	0.9844
LR			0.9232
ANN (9-46-1)	10%	0.9484	0.9708
LR			0.9092
ANN (9-46-1)	15%	0.9165	0.9493
LR			0.8870

ค่าเฉลี่ยผลลัพธ์ของการหาความต้องการน้ำอ้างอิงของพืชโดยที่มีค่าสูญหายพบว่า ค่าเฉลี่ยผลลัพธ์ R^2 ของวิธี ANN ยังให้ผลลัพธ์ที่ดีกว่า LR โดยที่เมื่อค่าสูญหายมีมากขึ้น ทำให้ค่าความแม่นยำมีค่าลดลงในทุกกรณี ดังรูปที่ 4.3



รูปที่ 4.3 แสดงค่าเฉลี่ยผลลัพธ์ของการหาความต้องการน้ำอ้างอิงของพืชโดยมีค่าสูญหาย ที่มีการทดแทนค่าสูญหายด้วยวิธี EM เปรียบเทียบระหว่าง วิธีการคำนวณแบบ ANN และ LR

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทดลองทดแทนค่าสูญหายด้วยการแทนค่าสูญหายจากวันใกล้ที่สุด โดยนำค่าในวันที่นั้นมาทดแทนค่าสูญหาย กำหนดค่า k ในวิธีการเป็นจำนวนวันแล้วหาค่าเฉลี่ย เท่ากับ 1 2 3 5 และ 10 แล้วทำการทดลองด้วยวิธีการคำนวณแบบ ANN และ LR เช่นเดียวกับการทดลองก่อนหน้านี้ได้ผลลัพธ์ดังตารางที่ 4.3

ตารางที่ 4.3 แสดงค่าเฉลี่ยผลลัพธ์ของการหาความต้องการน้ำอ้างอิงของพืชโดยมีค่าสูญหายที่มีการทดแทนค่าสูญหายด้วยวิธีแทนค่าสูญหายจากวันที่ใกล้ที่สุด เปรียบเทียบระหว่างวิธีการคำนวณแบบ ANN และ LR

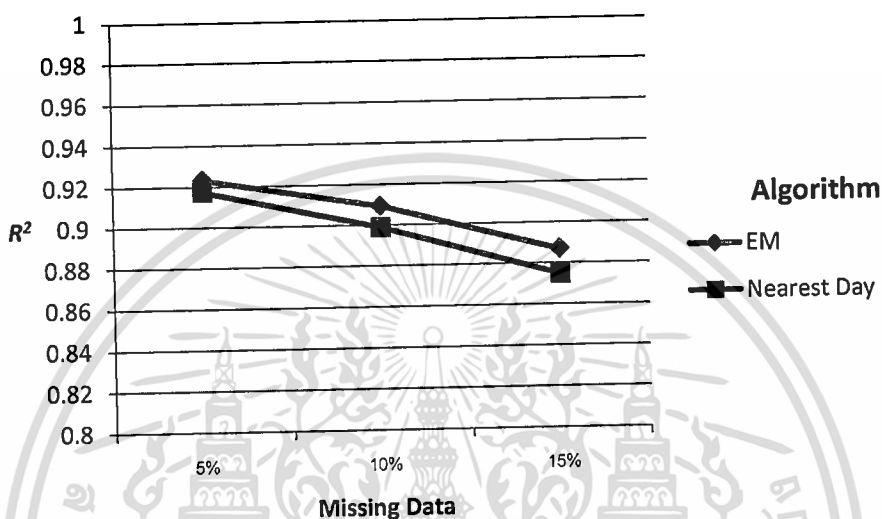


Algorithm	Percentage of Missing Values	Numbers of Nearest Day	R ² of Re-estimated Features	R ² of the Prediction
ANN (9-46-1)	5%	1	0.9789	0.9772
LR				0.9174
ANN (9-46-1)		2	0.9777	0.9748
LR				0.9152
ANN (9-46-1)		3	0.9763	0.9768
LR				0.9160
ANN (9-46-1)		5	0.9759	0.9771
LR				0.9164
ANN (9-46-1)	10	0.9756	0.9709	
LR			0.9168	
ANN (9-46-1)	10%	1	0.9599	0.9581
LR				0.8988
ANN (9-46-1)		2	0.9564	0.9567
LR				0.8982
ANN (9-46-1)		3	0.9571	0.9566
LR				0.8982
ANN (9-46-1)		5	0.9513	0.9552
LR				0.8972
ANN (9-46-1)	10	0.9496	0.9536	
LR			0.8980	
ANN (9-46-1)	15%	1	0.9339	0.9424
LR				0.8748
ANN (9-46-1)		2	0.9338	0.9424
LR				0.8722
ANN (9-46-1)		3	0.9328	0.9370
LR				0.8686
ANN (9-46-1)		5	0.9232	0.9245
LR				0.8672
ANN (9-46-1)	10	0.9210	0.9249	
LR			0.8704	

เอกสารนี้เป็นเอกสารทสวงนโสาหรับการใชงานเพือการศึกษาเท่านั้น ไม่นอญูาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

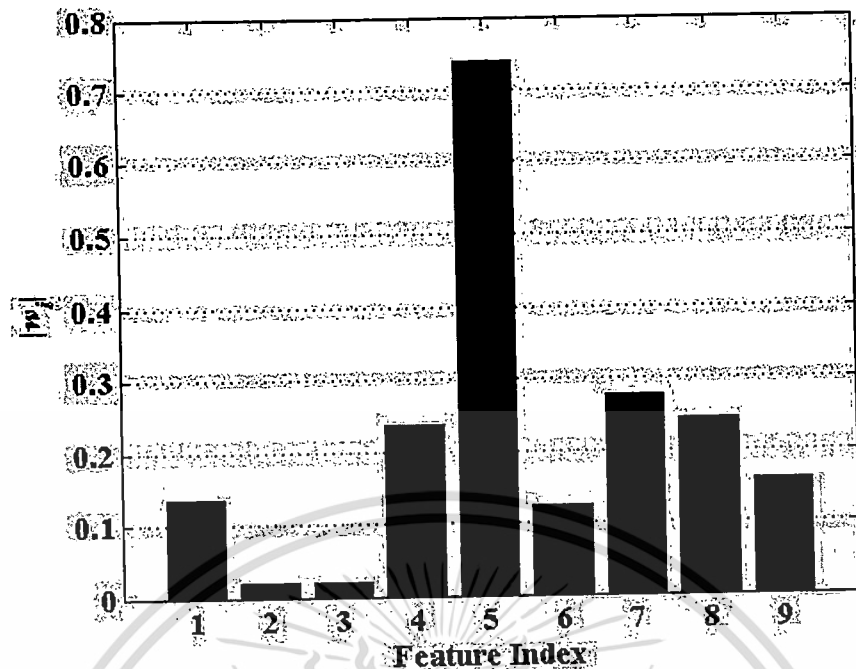
การทดลองทดแทนค่าสูญหายด้วยการแทนค่าสูญหายจากวันใกล้เคียงที่สุด โดยนำค่าในวันที่
 นั้นมาทดแทนค่าสูญหายพบว่า วิธีที่ดีที่สุดในการทดแทนค่าสูญหายจากวันที่ใกล้เคียงที่สุด ใช้วันเพียง 1
 วันจะทำให้ได้ผลลัพธ์ที่ดีกว่า ใช้เวลาหลายวันแล้วหาค่าเฉลี่ย สามารถเปรียบเทียบประสิทธิภาพใน
 การหาความต้องการน้ำอ้างอิงของพืช โดยมีค่าสูญหายด้วยวิธีเปรียบเทียบกับวิธี EM โดยใช้วิธี ANN
 ในการคำนวณ ปรากฏว่าวิธี EM มีประสิทธิภาพที่ดีกว่า ได้ผลลัพธ์ดังรูปที่ 4.4



รูปที่ 4.4 แสดงค่าเฉลี่ยผลลัพธ์ของการหาความต้องการน้ำอ้างอิงของพืช โดยมีค่าสูญหาย ที่มีการ
 ทดแทนค่าสูญหายด้วยวิธี EM และ การทดแทนค่าสูญหายจากวันที่ใกล้เคียงที่สุด โดยใช้วิธี
 ANN ในการคำนวณ

4.3 การหาความต้องการน้ำของพืชอ้างอิงโดยการเลือกใช้ตัวแปรลักษณะจำเพาะ

ในการศึกษาหาความสำคัญของตัวแปรลักษณะจำเพาะ ทดลองใช้ LR เพื่อค้นหา Linear
 Combination ของตัวแปรลักษณะจำเพาะที่ใช้ในการคำนวณหาความต้องการน้ำอ้างอิงของพืช ค่า
 น้ำหนักของ จะขึ้นอยู่กับความสัมพันธ์ของตัวแปรกับค่าผลลัพธ์ จากการทดลองที่ 1 เราสามารถ
 คำนวณค่าเฉลี่ยของค่าน้ำหนักจากการทดลอง LR ทั้ง 5 ครั้ง ซึ่งมีผลลัพธ์ตามรูปที่ 4.5



รูปที่ 4.5 แสดงค่าเฉลี่ย ความสำคัญของตัวแปรลักษณะจำเพาะ ในรูปแบบของค่าน้ำหนัก ($|w_i|$) LR ของทุกการทดลอง LR ในการทดลองที่ 1

ค่าน้ำหนักของตัวแปรลักษณะจำเพาะ แสดงให้เห็นว่าตัวแปรที่มีความสำคัญที่สุดคือ ความยาวนานแสงแดด เราสามารถเรียงลำดับความสำคัญของตัวแปรลักษณะจำเพาะได้ตามลำดับต่อไปนี้

$$5 > 7 > 8 > 4 > 9 > 1 > 6 > 3 \sim 2$$

จากค่าน้ำหนักของ LR ในรูปที่ 4.3 ได้ทำการทดลองเพื่อหาความต้องการน้ำอ้างอิงของพืช ด้วยวิธี ANN และ LR โดยเรียงลำดับตามตัวแปรลักษณะจำเพาะที่มีค่าความสำคัญมากที่สุด และเพิ่มขึ้นทีละ 1 ตัวในการทดลองครั้งต่อไปจนครบทุกตัว ตามลำดับความสำคัญของตัวแปรเรียงจากมากไปน้อย ได้ผลลัพธ์ดังตารางที่ 4.4

ตารางที่ 4.4 ผลการเปรียบเทียบประสิทธิภาพระหว่างการหาความต้องการน้ำอ้างอิงของพืชด้วยวิธี ANN และ LR เมื่อเลือกใช้ตัวแปรลักษณะจำเพาะ

Features Used	Algorithms	R^2 of the Prediction
5	ANN (1-10-1)	0.5889
	LR	0.5146
5,7	ANN (2-10-1)	0.8291
	LR	0.8105
5,7,8	ANN (3-41-1)	0.8833
	LR	0.8516
5,7,8,4	ANN (4-24-1)	0.9413
	LR	0.9126
5,6,8,4,9	ANN (5-32-1)	0.9586
	LR	0.9336
5,7,8,4,9,1	ANN (6-30-1)	0.9967
	LR	0.9361
5,7,8,4,9,1,6	ANN (7-34-1)	0.9958
	LR	0.9361
5,7,8,4,9,1,6,3	ANN (8-22-1)	0.9999
	LR	0.9371
5,7,8,4,9,1,6,3,2	ANN (9-46-1)	0.9999
	LR	0.9372

พบว่าวิธี ANN มีประสิทธิภาพในการหาความต้องการน้ำอ้างอิงของพืช ดีกว่าวิธี LR ในการใช้วิธี ANN เมื่อมีจำนวนตัวแปรลักษณะจำเพาะเพิ่มมากขึ้นจะทำให้มีประสิทธิภาพมากขึ้น เมื่อเพิ่มตัวแปรลักษณะจำเพาะตัวที่ 7 (อุณหภูมิต่ำสุด) จากที่ใช้ตัวแปรลักษณะจำเพาะตัวที่ 5 (ความยาวนานแสงแดด) เพียงตัวเดียวทำให้ประสิทธิภาพเพิ่มขึ้นถึง 40.78% เมื่อเพิ่มตัวแปรลักษณะจำเพาะ 8 (อุณหภูมิเฉลี่ย) และ 4 ความเร็วลมจะทำให้ประสิทธิภาพเพิ่มขึ้น 6.53% และ 6.56% เพิ่มตัวแปรลักษณะจำเพาะ 9 (ความชื้นสัมพัทธ์) และ 1 (วันที่ของปี) ทำให้ประสิทธิภาพเพิ่มขึ้นอีก 1.84% และ 3.97% แต่เมื่อเพิ่มตัวแปรลักษณะจำเพาะ 6 (อุณหภูมิสูงสุด) ทำให้ประสิทธิภาพลดลง เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มา -0.09% เพิ่มตัวแปรลักษณะจำเพาะ 3 (ละติจูด) ประสิทธิภาพเพิ่มขึ้นอีกเล็กน้อย 0.412% แต่เมื่อเพิ่มตัวแปรลักษณะจำเพาะ 2 (ความสูงจากระดับน้ำทะเล) จะมีค่าประสิทธิภาพเท่าเดิม

การใช้วิธี LR พบว่า การเพิ่มตัวแปรลักษณะจำเพาะ 7 (อุณหภูมิต่ำสุด) ทำให้มีประสิทธิภาพเพิ่มขึ้น 57.5% เพิ่มตัวแปรลักษณะจำเพาะ 8 (อุณหภูมิเฉลี่ย) และ 4 ความเร็วลม ประสิทธิภาพเพิ่มขึ้น 5.07% และ 7.16% เพิ่มตัวแปรลักษณะจำเพาะ 9 (ความชื้นสัมพัทธ์) และ 1 (วันที่ของปี) ทำให้ประสิทธิภาพเพิ่มขึ้นอีก 2.30% และ 0.26% เพิ่มตัวแปรลักษณะจำเพาะ 6 (อุณหภูมิสูงสุด) ประสิทธิภาพมีค่าไม่เปลี่ยนแปลง เพิ่มตัวแปรลักษณะจำเพาะ 3 (ละติจูด) ประสิทธิภาพเพิ่มขึ้น 0.11% และเมื่อเพิ่มตัวแปรลักษณะจำเพาะ 2 (ความสูงจากระดับน้ำทะเล) ประสิทธิภาพเพิ่มขึ้น 0.01% แสดงค่าประสิทธิภาพเปรียบเทียบระหว่าง ANN และ LR ได้ดังรูปที่ 4.6



รูปที่ 4.6 แสดงความสัมพันธ์ของการเพิ่มขึ้น/ลดลง ของความแม่นยำของการหาความต้องการน้ำ อ่างอิงของพืชด้วยวิธี ANN และ LR

ทดลองเพิ่มค่าสูญหายเข้าไปในข้อมูลชุดทดสอบ โดยเลือกใช้ตัวแปรลักษณะจำเพาะเพิ่มขึ้นทีละตัว โดยใช้วิธี ANN เพื่อหาประสิทธิภาพในการหาความต้องการน้ำของพืชอ่างอิงเมื่อมีค่าสูญหายได้ผลลัพธ์ดังตารางที่ 4.5

ตารางที่ 4.5 ผลการเปรียบเทียบประสิทธิภาพระหว่างการทำความต้องการน้ำอ้างอิงของพืชด้วยวิธี ANN เมื่อเลือกใช้ตัวแปรลักษณะจำเพาะ และมีค่าสูญหาย

Features Used	Algorithms	Percentage of Missing Values	R^2 of Re-estimated Features	R^2 of the Prediction
5	ANN (1-6-1)	5%	0.9619	0.5792
		10%	0.9619	0.5792
		15%	0.9619	0.5792
5,7	ANN (2-10-1)	5%	0.9593	0.7844
		10%	0.9184	0.7338
		15%	0.8718	0.6868
5,7,8	ANN (3-41-1)	5%	0.9593	0.8358
		10%	0.9183	0.8102
		15%	0.8718	0.7875
5,7,8,4	ANN (4-24-1)	5%	0.9522	0.9017
		10%	0.9075	0.8755
		15%	0.8624	0.8465
5,6,8,4,9	ANN (5-32-1)	5%	0.9599	0.9278
		10%	0.9217	0.9051
		15%	0.8860	0.8778
5,7,8,4,9,1	ANN (6-30-1)	5%	0.9599	0.9783
		10%	0.9232	0.9563
		15%	0.9336	0.9336
5,7,8,4,9,1,6	ANN (7-34-1)	5%	0.9726	0.9801
		10%	0.9438	0.9584
		15%	0.9137	0.9362
5,7,8,4,9,1,6,3	ANN (8-22-1)	5%	0.9730	0.9860
		10%	0.9444	0.9652
		15%	0.9138	0.9443
5,7,8,4,9,1,6,3,2	ANN (9-46-1)	5%	0.9730	0.9860
		10%	0.9444	0.9652
		15%	0.9138	0.9443

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.5 เมื่อเพิ่มตัวแปรลักษณะจำเพาะเข้าไปในวิธี ANN แล้ว เมื่อมีตัวแปรลักษณะจำเพาะมากขึ้น จะทำให้มีประสิทธิภาพมากขึ้น ยกเว้นเมื่อเพิ่มตัวแปร 2 (ความสูงจากระดับน้ำทะเล) ประสิทธิภาพในการหาความต้องการน้ำอ้างอิงของพืชมีค่าเท่าเดิม และเมื่อมีค่าสูญหายมากขึ้นจะทำให้ค่าประสิทธิภาพลดลง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปและวิจารณ์ผลการทดลอง

การหาความต้องการน้ำอ้างอิงของพืชด้วยวิธี Machine Learning โดยการใช้ ANN จากข้อมูลสภาพอากาศในประเทศไทยพบว่า เป็นวิธีที่มีสามารถใช้แทนวิธี PM^{2.5} ได้โดยมีประสิทธิภาพดีกว่าวิธี LR เพราะว่าวิธี LR เป็นการคำนวณแบบไม่เชิงเส้น

ในสถานการณ์ที่มีข้อมูลสภาพอากาศมีค่าสูญหาย สามารถใช้ EM เพื่อแก้ปัญหาค่าสูญหายที่เกิดขึ้น ซึ่งเป็นเรื่องปกติในการเก็บข้อมูลจริง ทำให้สามารถหาความต้องการน้ำอ้างอิงของพืชได้ แต่เมื่อมีค่าสูญหายเพิ่มมากขึ้นจะส่งผลกระทบต่อค่าคำนวณ โดยทำให้ประสิทธิภาพมีค่าลดลง

ตัวแปรลักษณะจำเพาะที่มีความสำคัญมากที่สุดในการหาความต้องการน้ำอ้างอิงของพืชคือ ความยาวนานแสงแดด อุณหภูมิ และความเร็วลม

เราอาจสามารถพัฒนาประสิทธิภาพด้วยการใช้วิธีไม่เชิงเส้นวิธีอื่นได้ เช่น Kernel Regression หรือ Support Vector Regression เพื่อเพิ่มประสิทธิภาพในการหาความต้องการน้ำอ้างอิงของพืช ในการทดลองมีการใช้ข้อมูลจากสถานีตรวจอากาศ 2 สถานี อาจจะทำให้ไม่เห็นถึงผลกระทบของที่ตั้งและความสูงจากระดับน้ำทะเล ในการหาความต้องการน้ำอ้างอิงของพืช ถ้าหากมีข้อมูลจากสถานีมากกว่านี้ อาจจะทำให้ตัวแปรลักษณะจำเพาะสองตัวนี้มีประสิทธิภาพเพิ่มขึ้นในการหาความต้องการน้ำอ้างอิงของพืชได้

บรรณานุกรม

ดร. ฤทัยชนก จริงจิตร. เจาะลึก “Smart Farmer” แค่แนวคิดใหม่ หรือจะพลิกโฉมเกษตรไทย.
สำนักงานนโยบายและยุทธศาสตร์การค้า สำนักนโยบายและยุทธศาสตร์การค้าสินค้า
เกษตร [5]

รศ. ดร. กัลยา วาณิชย์บัญชา. 2554. หลักสถิติ. กรุงเทพฯ. ศูนย์หนังสือแห่งจุฬาลงกรณ์มหาวิทยาลัย
[10]

Allen, R. Pereira, L. Raes, D. and Smith, M. 1998. **Crop evapotranspiration: Guidelines for
computing crop water requirements.** Rome: Food and Agricultural Organization of
the United Nations (FAO). [2]

Archana, C. and Shrivastava, R. K. 2010. “Reference Crop Evapotranspiration Estimation using
Artificial Neural Networks.” **International Journal of Engineering Science and
Technology.** Vol. 2(9):4205-4212 [8]

Benzaghta, M. Mohammed, T. Ghazali, A. and Soom, M. 2012. “Prediction of evaporation in
tropical climate using artificial neural network and climate based models.” **Scientific
Research and Essays.** 7(36):3133-3148 [4]

Flach, P. 2012. **Machine Learning : The Art and Science of Algorithms that Make Sense of
Data.** Cambridge [9]

Jain, S. K. Nayak, P. C. Sudheer, K. P. 2008. “Models for estimating evapotranspiration using
artificial neural networks, and their physical interpretation.” **HYDROLOGICAL
PROCESSES.** 22:2225-2234 [6]

Kumar, M. Raghuvanshi, N. Singh, R. Wallender, W. and Pruitt, W. 2002. “Estimating
evapotranspiration using artificial neural network.” **Journal of Irrigation and
Drainage Engineering.** 123:224-233 [3]

Nationnal Statistical Office. 2013. **Summary of the labor force survey in Thailand.**
Bangkok:Ministry of Information and Communication Technology. [1]

Wang, Y. Traore, S. Kerh, T. 2008. "Neural Network Approach for Estimating Reference Evapotranspiration from Limited Climatic Data in Burkina Faso." **WSEAS TRANSACTIONS on COMPUTER**. 6(7):704-713 [7]



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

บทความวิจัยที่ได้เผยแพร่และตีพิมพ์

บทความวิจัยที่ได้รับการเผยแพร่และตีพิมพ์ในวารสารวิชาการและการประชุมวิชาการระดับนานาชาติในการศึกษาอิสระฉบับนี้มีรายละเอียดดังต่อไปนี้

1. Pasupa K., Thamwiwatthana, E. (2013) Prediction of Reference Evapotranspiration with Missing Data in Thailand, In: Proceeding of 5th International Conference on Information Technology and Electrical Engineering (ICITEE'2013), 7-8 Oct 2013, Yogyakarta, Indonesia, pp 181-186.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Prediction of Reference Evapotranspiration with Missing Data in Thailand

Kitsuchart Pasupa

Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, Thailand
Email: kitsuchart@it.kmitl.ac.th

Ek Thamwiwatthana

Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, Thailand
Email: nupippo@gmail.com

Abstract—Artificial Neural Networks (ANNs) has been used in prediction of reference evapotranspiration for a recent decade. Its performance is competitive to a widely used method the so-called “Penman-Monteith” method. In this study, we aim to estimate the crop evapotranspiration by ANNs from climatic data in Thailand and compare the performance with the Penman-Monteith method. As missing data is inevitable, we also included the missing data situation into the study. This can be solved by expectation-maximization algorithm. The accuracy of the prediction decreases when the amount of missing values increases. Furthermore, we exploit the feature selection in the study. It shows that sunshine duration is the most important feature followed by temperature and wide speed, respectively.

Index Terms—reference evapotranspiration; missing data; neural network; feature selection

I. INTRODUCTION

Agriculture is an important mean of food production, and food is a basic necessity of human life. Nowadays, information and communication technology (ICT) begins to play a role in agriculture in order to reduce the cost and time of production. This is the so-called “precision farming”. It can influence the agricultural market competitiveness. Moreover, ICT can be utilized to water supply management in order to cope with drought and rain risks.

Thailand is an agricultural country. According to the recent summary of the labor force survey in Thailand reported by National Statistical Office, Ministry of Information and Communication Technology in March 2013, 35.15% of Thailand's labor force is employed in agriculture [1]. Moreover, Thai agriculture is very competitive and highly required water supply management, therefore precision farming is needed.

One of the approaches to water supply management is to schedule irrigation for high water use efficiency which can be done by estimating the crop evapotranspiration (ET_c). ET_c can be calculated from a multiplication between crop coefficient and reference evapotranspiration (ET_0). The value of ET_0 can be measure from Lysimeter. This method is very expensive and can only be utilized by well-trained person, therefore there are many approaches introduced to indirectly estimate the value of ET_0 namely Penman-Monteith equation [2], Hargreaves equation [3], etc. Penman-Monteith equation is one of the most globally used method for ET_0

estimation [2]. It can achieve the highest accuracy among the other methods when there is enough data.

Artificial intelligence techniques were recently applied to estimate the value of ET_0 e.g. artificial neural network [4], [5], [6], [7]. Artificial neural networks (ANNs) are trained from data collected by sensors which are installed in the crop fields. The data consists of pairs of input objects (e.g. temperature, humidity, solar radiation, wind speed) and desired outputs (i.e. ET_0). Once the model has trained, it can predict the value of ET_0 from the input data collected from the sensors. Evapotranspiration is very complex and nonlinear model. However, only a single hidden layer ANN is enough to mimic the model together with six inputs which are minimum and maximum temperature, minimum and maximum relative humidity, wind speed, and solar radiation [4]. ANN was also successfully applied to estimate evaporation rate with air temperature, humidity, wind velocity and solar radiation [5].

Problems arise when a size of the crop field is very large. The cost of the sensor installations will be increased according to the size and some area may not be able to install the devices. Therefore, many researchers aim to analyze and report of which input features to be used in the prediction. Hence the cost and the number of devices could be reduced. Moreover, the computational time is reduced. There is an evidence that using inputs of air temperature, wind speed, humidity and solar radiation in Malaysia gives the highest accuracy of the prediction [5]. However, the selection of variables depends on the area where the experiment is conducted such as USA [4] Malaysia [5], Brazil [6], and Burkina Faso [7]. To the best of the authors' knowledge, there is no report about this matter in Thailand.

In real-world implementation, data loss is inevitable. Missing data can occur for many reasons: storage or sensor mechanisms are malfunction, system fails to respond when the data is transmitted. Therefore, it is not able to estimate the current reference evapotranspiration. We can use classic approaches to solve this problem. This can be done by mean substitution or list-wise deletion. However, these approaches can reduce the prediction accuracy. One of the popular statistical methods used for estimating missing values is Expectation-Maximization (EM) algorithm. It aims to compute the maximum likelihood estimation in the presence of missing data [8].

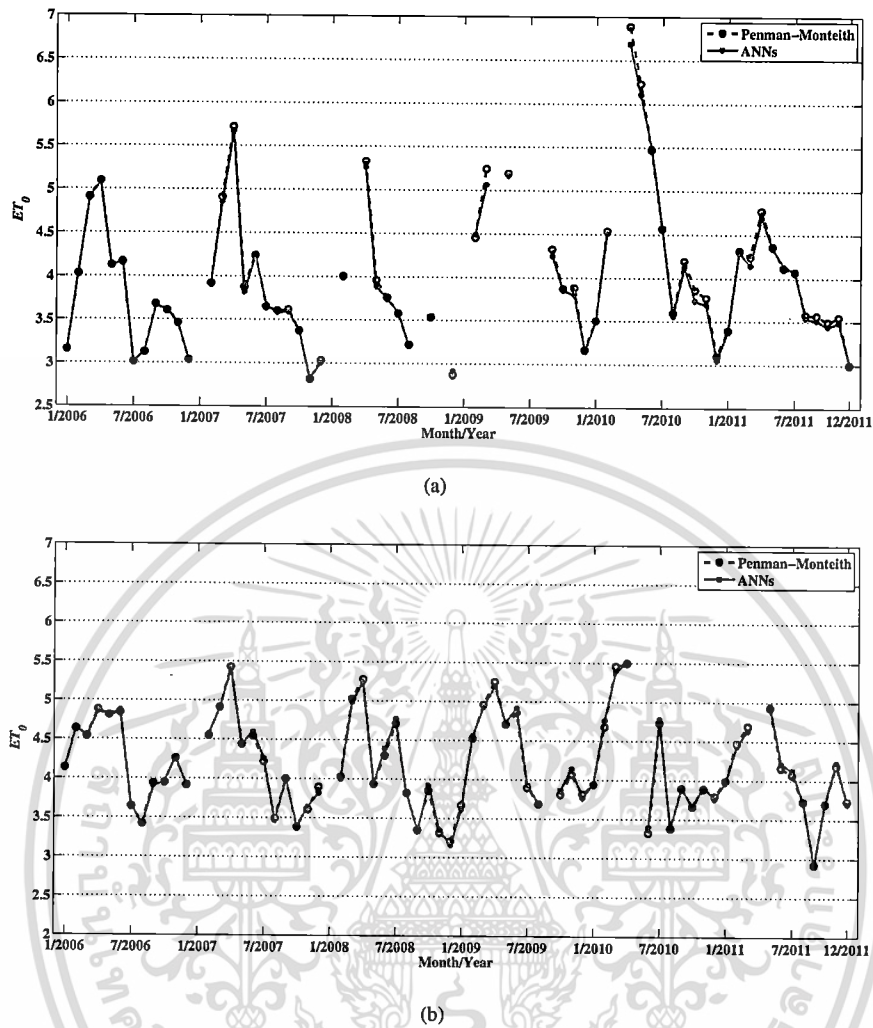


Fig. 3. Monthly reference evapotranspiration from 2006 to 2011 in (a) Ubon Ratchathani province and (b) Chiang Mai province by Penman-Monteith equation and ANNs.

models. Then we calculated the average of the absolute values of each element in w_i across five vectors. According to Fig. 5, it can be seen that sunshine duration is the most important feature. Moreover, we ranked the values of absolute weight in descending order which gives the following order:

$$5 > 7 > 8 > 4 > 9 > 1 > 6 > 3 \sim 2$$

In this scenario, we trained models by adding new features one-by-one according to how important the features are. Five-fold cross validation is used to evaluate the performances and search for the optimal parameters. Missing value cases were also considered too.

Table IV shows the performances of ANN and LR when different set of features are used. The classifiers with 8 features (i.e. feature 5, 7, 8, 4, 9, 1, 6, and 3) and 9 features gave the highest accuracy for both ANN and LR cases. Fig. 6 shows the relative improvement or worsening in the accuracy for ANN and LR when we considered new features in the

process. In ANN, adding feature 7 (minimum temperature) to feature 5 (sunshine duration) could improve the performance by 40.99% of using only feature 5. In addition, when we considered mean temperature and wind speed in the process, the performances were improved 5.7% and 7.2%, respectively. Again, the performance were improved by 1.6% and 4.2% after adding relative humidity and day-of-year, respectively. Unfortunately, considering maximum temperature, latitude, and high above mean sea level in addition, did not improve much. Overall picture is much the same for LR.

Moreover, we also examined the feature selection process together with missing values case by ANN as shown in Table V. The overall picture is still much the same for the case with no missing values. The performances of the algorithms drop when the amount of missing values increases. Here, ANNs (8-49-1) with 5% missing values yield the best performance.

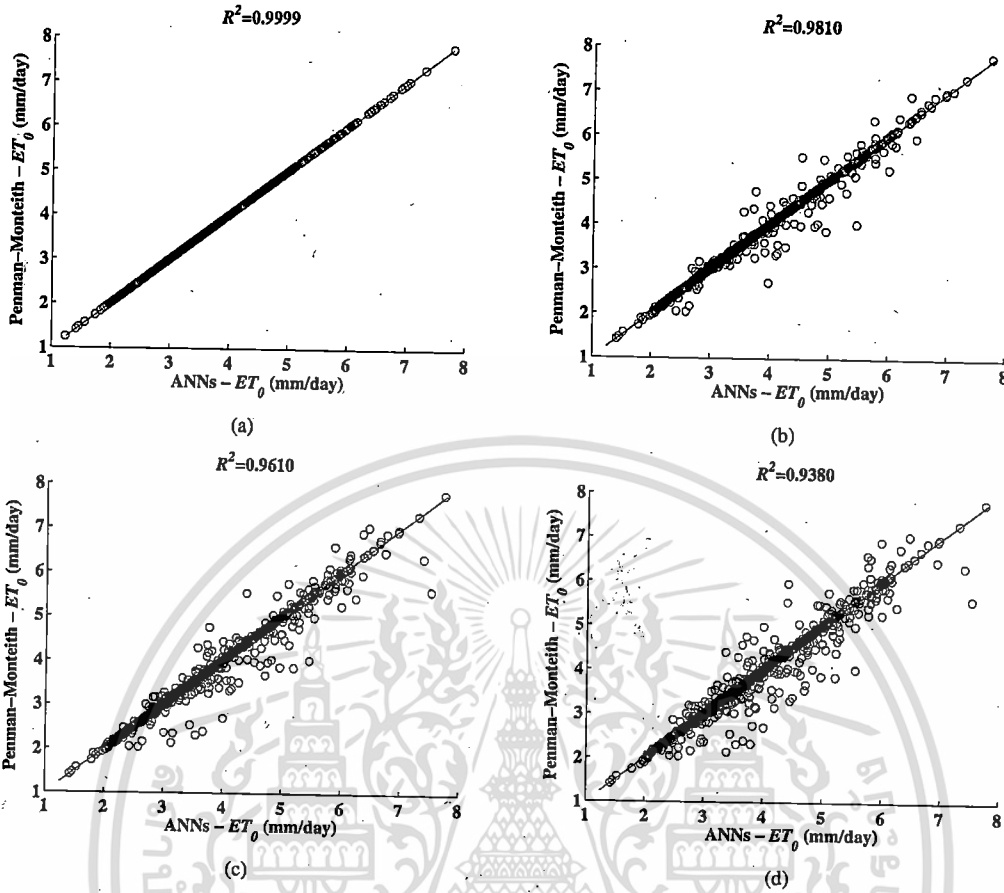


Fig. 4. Comparison of ET_0 computed by Penman-Monteith equation with ANNs (9-38-1) model when there is (a) no missing value, (b) 5% of missing value, (c) 10% of missing value and (d) 15% of missing value.

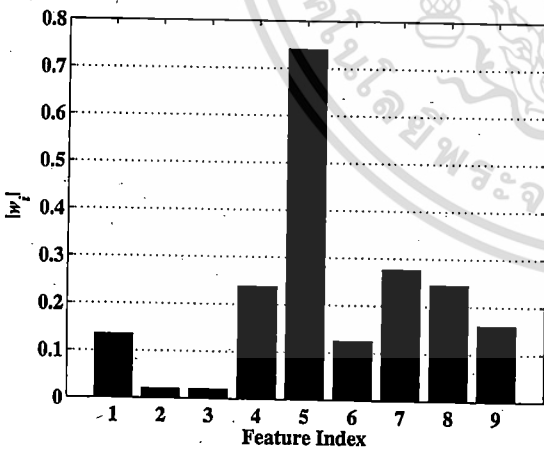


Fig. 5. An illustration of feature importance. It shows the absolute value of the linear regression weight vectors w_i average across all the models.

TABLE IV
PERFORMANCE COMPARISON BETWEEN ANNS AND LR WHEN FEATURE SELECTION IS CONSIDERED.

Features used	Algorithms	R^2
5	ANN (1-6-1)	0.5889
	LR	0.5709
5,7	ANN (2-10-1)	0.8303
	LR	0.8010
5,7,8	ANN (3-12-1)	0.8776
	LR	0.8511
5,7,8,4	ANN (4-32-1)	0.9410
	LR	0.9122
5,7,8,4,9	ANN(5-21-1)	0.9561
	LR	0.9260
5,7,8,4,9,1	ANN(6-29-1)	0.9963
	LR	0.9355
5,7,8,4,9,1,6	ANN(7-40-1)	0.9960
	LR	0.9355
5,7,8,4,9,1,6,3	ANN(8-49-1)	0.9999
	LR	0.9365
5,7,8,4,9,1,6,3,2	ANN(9-38-1)	0.9999
	LR	0.9365

IV. CONCLUSIONS

In this paper, we used ANN and LR to predict the reference evapotranspiration from the climatic data in Thailand. We have shown that ANN is competitive to the most widely

used method the so-called "Penman-Monteith equation". The performance of LR is generally worse than ANN, however, this could be improved by using kernel regression as the evapotranspiration is nonlinear model. We also applied EM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

TABLE V
PERFORMANCE OF ANNs WHEN FEATURE SELECTION IS CONSIDERED TOGETHER WITH MISSING VALUE SITUATIONS.

Features used	Algorithms	Percentage of missing values	R^2 of re-estimated features	R^2 of the prediction
5	ANN (1-6-1)	5%	0.9619	0.5752
		10%	0.9619	0.5752
		15%	0.9619	0.5752
5,7	ANN (2-10-1)	5%	0.9586	0.8046
		10%	0.9149	0.7610
		15%	0.8576	0.7140
5,7,8	ANN (3-12-1)	5%	0.9590	0.8544
		10%	0.9186	0.8306
		15%	0.8665	0.8070
5,7,8,4	ANN (4-32-1)	5%	0.9543	0.9086
		10%	0.9090	0.8803
		15%	0.8598	0.8520
5,7,8,4,9	ANN (5-21-1)	5%	0.9619	0.9329
		10%	0.9231	0.9098
		15%	0.8860	0.8838
5,7,8,4,9,1	ANN (6-29-1)	5%	0.9619	0.9772
		10%	0.9231	0.9527
		15%	0.8809	0.9272
5,7,8,4,9,1,6	ANN (7-40-1)	5%	0.9740	0.9772
		10%	0.9456	0.9674
		15%	0.9112	0.9335
5,7,8,4,9,1,6,3	ANN (8-49-1)	5%	0.9740	0.9818
		10%	0.9456	0.9605
		15%	0.9112	0.9390
5,7,8,4,9,1,6,3,2	ANN (9-38-1)	5%	0.9740	0.9835
		10%	0.9456	0.9617
		15%	0.9112	0.9391

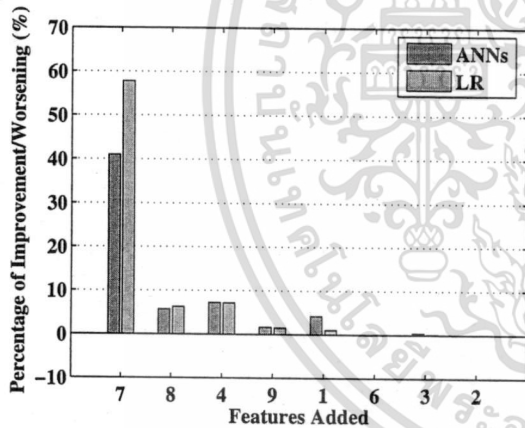


Fig. 6. Relative improvement/worsening in accuracy of ANN and LR when adding new features one-by-one according to how important the features are.

algorithm to solve the problem when missing data occur in real-world implementation. Therefore, we still can predict the reference evapotranspiration. When the amount of missing data increases, the performances of the re-estimated feature and the prediction of the learning algorithms will drop. However, this is still worthwhile. Moreover, the feature selection process was performed. We have shown the importance of the features used in the prediction. The most three importance features are sunshine duration, temperature and wind speed, respectively. In this study, we have only investigated on two automatic weather stations, therefore, we did not see much impact to the prediction with the height above mean sea level and latitude

of the stations. However, if there is a number of automatic weather stations, these two features might have more effect to the prediction.

REFERENCES

- [1] National Statistical Office, "Summary of the labor force survey in Thailand: March 2013," Ministry of Information and Communication Technology, Bangkok, Thailand, 2013.
- [2] R. Allen, L. Pereira, D. Raes, and M. Smith, "Crop evapotranspiration-guidelines for computing crop water requirements," *FAO Irrigation and Drainage Paper*, vol. 56, pp. 15-78, 1998.
- [3] G. Hargreaves and R. Allen, "History and evaluation of hargreaves evapotranspiration equation," *Journal of Irrigation and Drainage Engineering*, vol. 129, no. 1, pp. 53-63, 2003.
- [4] M. Kumar, N. Raghuvanshi, R. Singh, W. Wallender, and W. Pruitt, "Estimating evapotranspiration using artificial neural network," *Journal of Irrigation and Drainage Engineering*, vol. 128, no. 4, pp. 224-233, 2002.
- [5] M. Benzaghta, T. Mohammed, A. Ghazali, and M. Soom, "Prediction of evaporation in tropical climate using artificial neural network and climate based models," *Scientific Research and Essays*, vol. 7, no. 36, pp. 3133-3148, 2012.
- [6] S. Zanetti, E. Sousa, V. Oliveira, F. Almeida, and S. Bernardo, "Estimating evapotranspiration using artificial neural network and minimum climatological data," *Journal of Irrigation and Drainage Engineering*, vol. 133, no. 2, pp. 88-89, 2012.
- [7] T. Wang, S. Traore, and T. Kerh, "Neural network approach for estimation reference evapotranspiration from limited climatic data in Burkina Faso," *WSEAS Transactions on Computers*, vol. 7, no. 6, pp. 704-713, 2008.
- [8] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1-38, 1977.
- [9] A. Dongare, R. Kharde, and A. Kachare, "Introduction to artificial neural network," *International Journal of Engineering and Innovative Technology*, vol. 2, no. 1, pp. 189-194, 2012.
- [10] Royal Command of H.M. the King, "Climate," *Thai Junior Encyclopedia*, vol. 4, no. 6, 1978.