บทที่ 4
การพัฒนาแบบจำลองกระแสน้ำในอ่าวไทยโดยใช้ซัพพอร์ตเวกเตอร์แมทชีน

ในบทนี้จะกล่าวถึงการลำดับขั้นตอนการพัฒนาแบบจำลองกระแสน้ำในอ่าวไทย การเตรียมและเก็บรวบรวมข้อมูลสำหรับใช้สอนซัพพอร์ตเวกเตอร์แมทชีน เพื่อให้ทำหน้าที่เป็นแบบจำลองกระแสน้ำ และเทคนิคการคัดเลือกแบบจำลองที่เหมาะสมที่สุด

4.1 วิธีการพัฒนาแบบจำลองกระแสน้ำโดยใช้ซัพพอร์ตเวกเตอร์แมทชีน

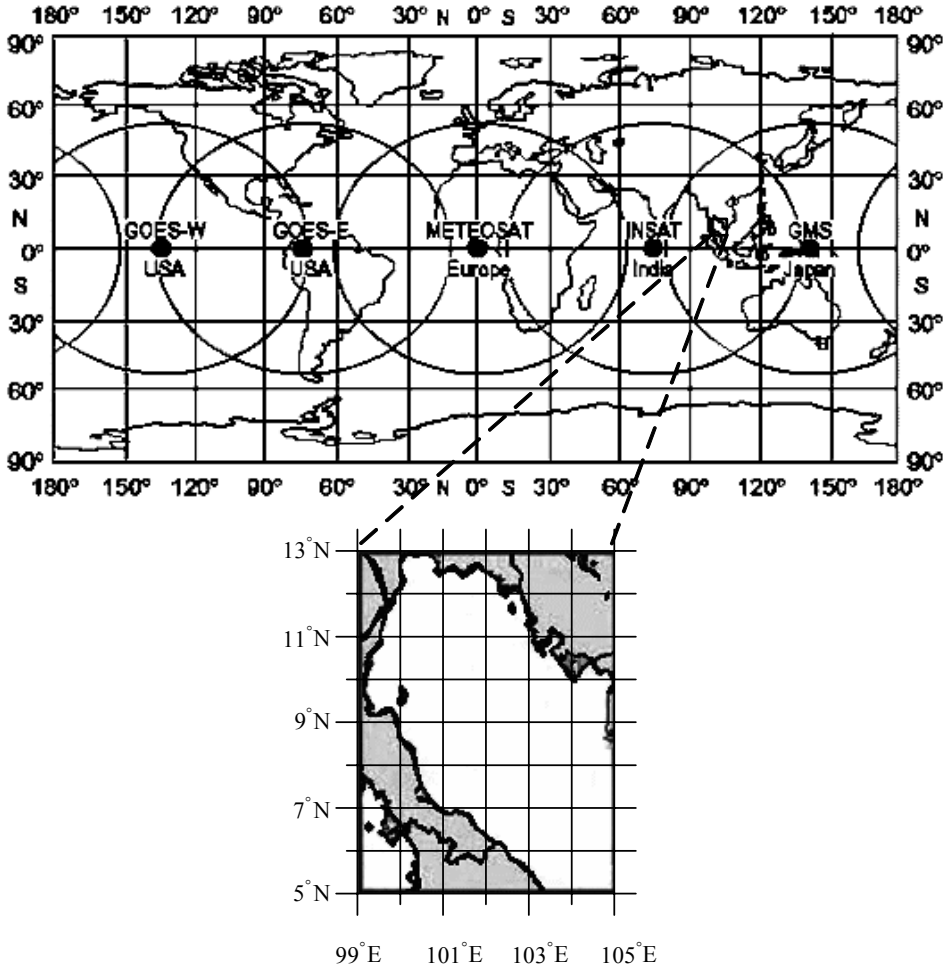
งานวิจัยนี้ได้ทำการพัฒนาและออกแบบแบบจำลองกระแสน้ำในอ่าวไทย ที่มีการทำงานเลียนแบบมาจากการทำงานของแบบจำลองกระแสน้ำพรินซ์ตัน โดยได้เลือกใช้ซัพพอร์ตเวกเตอร์แมทชีนสำหรับการถดถอย (Support Vector Regression: SVR) เนื่องจากงานวิจัยนี้เป็นงานที่เกี่ยวข้องกับการประมาณค่าฟังก์ชัน อีกทั้งซัพพอร์ตเวกเตอร์แมทชีนมีสมรรถนะโดยทั่วไปที่เหนือกว่าโครงข่ายประสาทเทียมแบบดั้งเดิมตามที่ได้กล่าวไว้ในบทที่ผ่านมา

สำหรับการทำงานของแบบจำลองกระแสน้ำพรินซ์ตันนั้น หลังจากที่กำหนดค่าพารามิเตอร์เริ่มต้นให้ใช้งานกับทะเลไทยได้แล้ว ในขั้นตอนของการใช้งานจริงจะใช้เพียงข้อมูลที่มีการเปลี่ยนแปลงซึ่งได้แก่ ข้อมูลลม กับตำแหน่งเริ่มต้นของวัตถุเท่านั้น ขั้นตอนของการเรียนรู้ของซัพพอร์ตเวกเตอร์แมทชีนจะอาศัยข้อมูล อินพุต-เอาต์พุต จากแบบจำลองกระแสน้ำพรินซ์ตันเป็นข้อมูลต้นแบบ ข้อมูลอินพุตจะสนใจเฉพาะข้อมูลที่มีการเปลี่ยนแปลงเท่านั้น ส่วนข้อมูลเอาต์พุตจะเหมือนกันกับแบบจำลองกระแสน้ำพรินซ์ตัน ขั้นตอนการทำงานอย่างง่ายแสดงดังรูป 4.1

รูป 4.1 แบบจำลองกระแสน้ำที่ใช้ซัพพอร์ตเวกเตอร์แมทชีนสำหรับการถดถอย

4.2 การเตรียมข้อมูลสำหรับฝึกสอนซัพพอร์ตเวกเตอร์แมทชีน

เซตของข้อมูลอินพุตที่เตรียมสำหรับฝึกสอนซัพพอร์ตเวกเตอร์แมทชีนนั้น จะใช้ข้อมูลชุดเดียวกันกับที่ป้อนให้กับแบบจำลองกระแสน้ำพรินซ์ตัน คือ ข้อมูลลมที่ความสูง 10 เมตรเหนือผิวหน้าน้ำทะเลจากเว็บไซต์ของห้องสมุดสิ่งแวดล้อมมาสเตอร์ (URL: http://mel.dmso.mil) ข้อมูลที่ได้รับจะมีความละเอียดทุกๆ หนึ่งองศาของจุดที่ตัดกันระหว่างเส้นละติจูดกับเส้นลองจิจูดทั่วโลก ซึ่งมีข้อมูลทั้งหมดเท่ากับ 181×360 ข้อมูล แสดงรายละเอียดดังรูป 4.2 แต่แบบจำลองกระแสน้ำพรินซ์ตันจะอ่านข้อมูลเฉพาะบริเวณพื้นที่ที่ได้กำหนดการทำงานไว้ สำหรับงานวิจัยนี้ได้มีการกำหนดพื้นที่ทำการวิจัยเฉพาะในบริเวณทะเลฝั่งอ่าวไทยอาณาเขตตั้งแต่ เส้นละติจูดที่ 5-13 องศาเหนือ เส้นลองจิจูดที่ 99-105 องศาตะวันออก ดังนั้นจึงต้องมีการปรับพื้นที่การทำงานของแบบจำลองกระแสน้ำพรินซ์ตันใหม่ และต้องลดขนาดข้อมูลลงเพื่อใช้กับซัพพอร์ตเวกเตอร์แมทชีนด้วยซึ่งจะกล่าวรายละเอียดในหัวข้อต่อไป



รูป 4.2 บริเวณพื้นที่อ่าวไทยเส้นละติจูดที่ 5-13°เหนือ เส้นลองจิจูดที่ 99-105°ตะวันออก

4.2.1 รายละเอียดของข้อมูลลม

Gridded Binary Data (GRIB) คือ รูปแบบมาตรฐานของการเข้ารหัสสำหรับจัดเก็บไฟล์ข้อมูลแบบไบนารีอีกรูปแบบหนึ่ง ที่ได้ถูกกำหนดขึ้นโดยองค์กรอุตุนิยมวิทยาโลก (World Meteorological Organization: WMO) [3,6] เนื่องจากข้อมูลที่เกี่ยวข้องกับทางด้านอุตุนิยมวิทยาจะมีขนาดข้อมูลที่มีขนาดใหญ่มาก เนื่องจากให้ความละเอียดของข้อมูลทุกๆ หนึ่งตารางกริดทั่วโลก หรือทุกๆ หนึ่งองศาของจุดตัดเส้นละติจูดกับเส้นลองจิจูด จึงทำให้ข้อมูลมีขนาดใหญ่ถึง 181×360 ไฟล์ข้อมูล (ละติจูดเหนือ 90 จุด ศูนย์สูตร 1 จุด ละติจูดใต้ 90 จุด ลองจิจูดตะวันออก 180 จุด และลองจิจูดตะวันตก 180 จุด) จึงจำเป็นต้องมีการกำหนดมาตรฐานรูปแบบของการจัดเก็บข้อมูล ให้มีความเข้าใจที่ตรงกัน เพื่อความสะดวกในการนำไปใช้งานในด้านอื่นๆ ที่เกี่ยวข้อง

เทคนิคพิเศษได้ถูกออกแบบสำหรับการจัดเก็บข้อมูล เพื่อให้เกิดความสะดวกในการนำไปใช้งานได้อย่างรวดเร็ว รวมไปถึงความง่ายต่อการส่งข้อมูลผ่านทางอินเตอร์เน็ตจากห้องสมุดสิ่งแวดล้อมมาสเตอร์ ซึ่งเป็นองค์กรที่ให้บริการข้อมูลทางด้านอุตุนิยมวิทยากับสมาชิกทั่วโลก ทั้งนี้การจัดเก็บข้อมูลในรูปแบบของกริบไฟล์นั้น จะกำหนดให้ไฟล์มีนามสกุลเป็น .grib มีการกำหนดรูปแบบที่เฉพาะตัว

ข้อมูลที่อยู่ในรูปแบบของกริบไฟล์จะเป็นข้อมูลที่เรียงอนุกรมกัน โดยจะประกอบไปด้วยส่วนหัวของข้อมูล และตามด้วยแพ็กเกจของข้อมูลซึ่งแทนตำแหน่งตำบลที่ 1 กริดใดๆ บนพื้นโลกด้วยข้อมูลขนาด 2 มิติ รายละเอียดในส่วนหัวของข้อมูล แสดงรายละเอียดที่เกี่ยวข้องเฉพาะกับชุดของข้อมูลนั้นๆ เช่น ค่าพารามิเตอร์ หน่วยความถูกต้องของข้อมูล ระบบการจัดเรียงกริดระดับความสูง กรรมวิธีของการผลิตข้อมูล ซึ่งความถูกต้องของข้อมูลจะขึ้นอยู่กับกรรมวิธีการผลิตข้อมูลนั้นๆ โดยจะแบ่งกรรมวิธีการผลิตข้อมูลออกเป็นข้อมูลพยากรณ์ล่วงหน้า (Forecast) และข้อมูลของช่วงเวลาที่ผ่านมาแล้ว (Hindcast) ซึ่งข้อมูลชนิดหลังนี้จะให้ความถูกต้องมากกว่า และนอกจากนี้ข้อมูลส่วนหัวยังบรรจุวันที่ และเวลาที่ผลิตข้อมูลนั้น รวมไปถึงการกำหนดส่วนเพิ่มเติมของผู้ให้ข้อมูล ซึ่งการเปลี่ยนแปลงรูปแบบที่มาตรฐานมีผลจะย้อนไปถึงความเข้ากันได้กับโปรแกรมที่ใช้ในการอ่านกริบไฟล์ด้วย ดังนั้นข้อมูลที่อยู่ในรูปแบบมาตรฐานของกริบไฟล์จะสามารถเข้าได้กับโปรแกรมมาตรฐานใดๆ ที่ใช้สำหรับอ่านกริบไฟล์ได้ทั้งสิ้น

ตัวอย่างของลักษณะไฟล์ที่ขอผ่านอินเตอร์เน็ต

058.240.wnd_ucmp.ht_sfc.10.0.2002082900.000.GRIB

058.240.wnd_vcmp.ht_sfc.10.0.2002083012.000.GRIB

ความหมายในส่วนของชื่อไฟล์

058.240	หมายถึง	รหัสข้อมูลตามมาตรฐาน IEEE
wnd_ucmp	หมายถึง	ลมในแนวแกนนอน
wnd_vcmp	หมายถึง	ลมในแนวแกนตั้ง
ht_sfc	หมายถึง	เป็นข้อมูลที่ผ่านมาแล้ว (Hindcast)
ft_sfc	หมายถึง	เป็นข้อมูลพยากรณ์ล่วงหน้า (Forecast)
10.0	หมายถึง	ข้อมูลที่ความสูง 10 เมตร
20020829	หมายถึง	ค.ศ./เดือน/วันที่
12.000	หมายถึง	เวลาที่ให้ข้อมูลตามเวลามาตรฐานกรีนิช (GMT)
GRIB	หมายถึง	การเข้ารหัสข้อมูลตามมาตรฐาน WMO

4.2.2 การถอดรหัสข้อมูลนามสกุลกริบ (*.grib)

สำหรับในงานวิจัยนี้จะใช้โปรแกรม Gridded Binary (GRIB) Software ซึ่งสามารถดาวน์โหลดได้จากเว็บไซต์ของ WMO สำหรับใช้อ่านข้อมูลลมที่ขอจากห้องสมุดสิ่งแวดล้อมมาสเตอร์ โดยการใช้คำสั่ง “wgrib” ซึ่งมีรูปแบบของการใช้งานดังนี้

>> wgrib -d all -text -o [filename].text [filename].grib

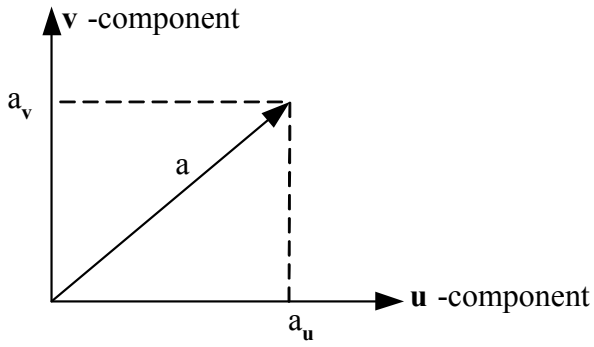
หมายถึง ให้ทำการถอดรหัสข้อมูลจากโครงสร้างแบบกริบเปลี่ยนเป็นโครงสร้างข้อมูลแบบเท็กซ์ไฟล์ ซึ่งงานวิจัยนี้จะใช้ข้อมูลแบบเท็กซ์ไฟล์ในการฝึกสอนซัพพอร์ตเวกเตอร์แมทชีน สำหรับข้อมูลที่ถอดรหัสออกมาได้นั้นจะเป็นข้อมูลลมทุกๆ กริดทั่วโลกเรียงลำดับต่อกันขนาด 65,160×1 ดังตัวอย่าง

```
 360 181
-1.59303
 1.69303
-1.69303
-1.79303
-1.89303
 1.89303
 1.99303
 1.99303
-2.09303
    :
    :
```

ข้อมูลบรรทัดแรกหมายถึงขนาดของข้อมูล ส่วนบรรทัดต่อๆ มาคือข้อมูลลมมีหน่วยเป็น เมตร/วินาที เครื่องหมายลบ หมายถึง ทิศทางลมมีทิศตรงข้ามกับทิศทางที่กำหนดไว้ เพื่อนำข้อมูลดังกล่าวมาใช้กับงานวิจัยนี้ จึงต้องนำข้อมูลมาวิเคราะห์เพื่อศึกษาถึงการจัดเรียงลำดับของข้อมูล และเลือกเอาเฉพาะข้อมูลช่วงบริเวณอ่าวไทย ซึ่งจะกล่าวถึงรายละเอียดในหัวข้อต่อไป

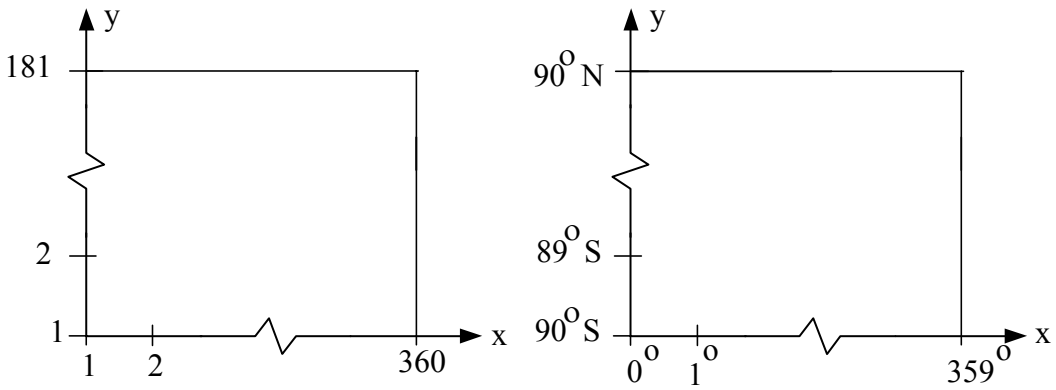
4.2.3 รูปแบบของการจัดเรียงข้อมูลทั่วโลก

ข้อมูลลมที่ขอจากห้องสมุดสิ่งแวดล้อมมาสเตอร์ จะมีการจัดรูปแบบของข้อมูลในลักษณะองค์ประกอบของเวกเตอร์แนวนอน (**u**-component) และเวกเตอร์แนวดิ่ง (**v**-component) กล่าวคือ ข้อมูลลมของแต่ละกริด แต่ละช่วงเวลาจะถูกแบ่งออกเป็น 2 ชุดใหญ่ๆด้วยกันได้แก่ ชุด **u** หมายถึง ลมที่พัดในทิศแนวนอน (ตะวันออก-ตะวันตก) และชุด **v** หมาย ถึงลมที่พัดในทิศแนวดิ่ง (เหนือ-ใต้) แสดงดังรูป 4.3



รูป 4.3 ลักษณะของข้อมูลลมในระบบกริบ

การจัดเรียงข้อมูลในระบบกริบ จะเป็นการนำข้อมูลของแต่ละกริดมาเรียงลำดับต่อกันไปตามแนวแกน x และ y รอบโลก โดยกำหนดตำแหน่งข้อมูลเริ่มต้น (1,1) จะตรงกับตำแหน่งเส้นลองจิจูดที่ 0° เส้นละติจูดที่ 90° ใต้ โดยจะเรียงลำดับข้อมูลในแนวนอนจากซ้ายไปขวา และเรียงลำดับข้อมูลในแนวตั้งจากล่างขึ้นบนดังรูป 4.4 ดังนั้นตำแหน่งข้อมูลลำดับสุดท้าย (ลำดับที่ 65,160) จะตรงกับคอลัมน์ที่ 360 แถวที่ 181 ซึ่งจะตรงกับตำแหน่งเส้นลองจิจูดที่ 359° หรือ 179° ตะวันตก เส้นละติจูดที่ 90° เหนือ

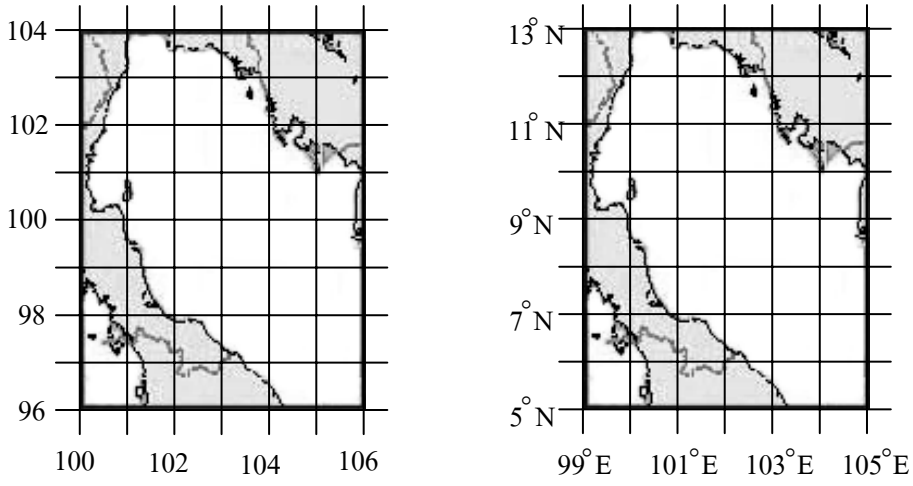


รูป 4.4 การจัดเรียงข้อมูลในระบบกริบ

จากข้อมูลที่ได้รับทั่วโลกขนาดมิติเท่ากับ 181×360 นำมาวิเคราะห์ข้อมูลเฉพาะในส่วนของบริเวณอ่าวไทย ซึ่งมีอาณาเขตตั้งแต่ เส้นลองจิจูดที่ 99-105°ตะวันออก เส้นละติจูดที่ 5-13°เหนือ ทำให้ขนาดมิติของข้อมูลลดลงเหลือเท่ากับ 9×7 ส่วนกรรมวิธีในการลดขนาดของข้อมูลจะกล่าวรายละเอียดในหัวข้อถัดไป

4.2.4 การกำหนดพื้นที่ศึกษาข้อมูลเพื่อใช้กับงานวิจัย

จากข้อมูลข้างต้นเมื่อเทียบกับข้อมูลบริเวณตำแหน่งพื้นที่ที่จะทำการศึกษาในงานวิจัยนี้จะอยู่ในช่วงตั้งแต่คอลัมน์ที่ 100-106 และอยู่ในช่วงตั้งแต่แถวที่ 96-104 ซึ่งจะตรงกับลำดับข้อมูลที่ 34,301-34,307; 34,661-34,667; 35,021-35,027; 35,381-35,387; 35,741-35,747; 36,101-36,107; 36,464-36,467; 36,821-36,827 และ 37,181-37,187 ตามลำดับตามรูป 4.5 และดูรายละเอียดในตาราง 4.1



รูป 4.5 ตำแหน่งข้อมูลบริเวณพื้นทำงานวิจัย

ตาราง 4.1 รายละเอียดของตำแหน่งข้อมูลบริเวณพื้นที่ทำงานวิจัย

Row\Col	100	101	102	103	104	105	106
104	37,181	37,182	37,183	37,184	37,185	37,186	37,187
103	36,821	36,822	36,823	36,824	36,825	36,826	36,827
102	36,461	36,462	36,463	36,464	36,465	36,466	36,467
101	36,101	36,102	36,103	36,104	36,105	36,106	36,107
100	35,741	35,742	35,743	35,744	35,745	35,746	35,747
99	35,381	35,382	35,383	35,384	35,385	35,386	35,387
98	35,021	35,022	35,023	35,024	35,025	35,026	35,027
97	34,661	34,662	34,663	34,664	34,665	34,666	34,667
96	34,301	34,302	34,303	34,304	34,305	34,306	34,307

เมื่อทราบตำแหน่งของข้อมูลที่ต้องการจะศึกษาแล้ว ก็สามารถนำไปออกแบบโปรแกรมสำหรับการเลือกช่วงข้อมูลให้ตรงกับตำแหน่งที่ต้องการจะศึกษาต่อไป

4.2.5 รายละเอียดการเตรียมข้อมูลลมสำหรับใช้สอนซัพพอร์ตเวกเตอร์แมทชีน

สำหรับงานวิจัยนี้ใช้ข้อมูลลมย้อนหลัง (Hindcast) เนื่องจากให้ความถูกต้องของข้อมูลมากกว่าข้อมูลลมพยากรณ์ (Forecast) โดยมีรายละเอียดของวิธีการเก็บข้อมูลดังนี้

(1) ใช้ข้อมูลลมของปี ค.ศ. 2002 ตั้งแต่เดือนพฤษภาคม ถึง เดือนกรกฎาคม ซึ่งเป็นช่วงที่อ่าวไทยมีคลื่นลมแรง

(2) ในแต่ละเดือนสุ่มเลือกเก็บข้อมูล 8 วัน ใน 1 วัน ข้อมูลลมจะห่างกัน 12 ชั่วโมง ดังนั้นจึงเก็บข้อมูลได้ 2 ชุด/วัน

(3) ในแต่ละเดือนจะเก็บข้อมูลลมตัวอย่างได้ 16 ชุด ดังนั้นใน 3 เดือนจะมีข้อมูลทั้งหมด 48 ชุด

(4) ข้อมูลทั้ง 48 ชุดแยกออกเป็นข้อมูลลมในแนวแกน **u** และ **v** อย่างละ 48 ชุด

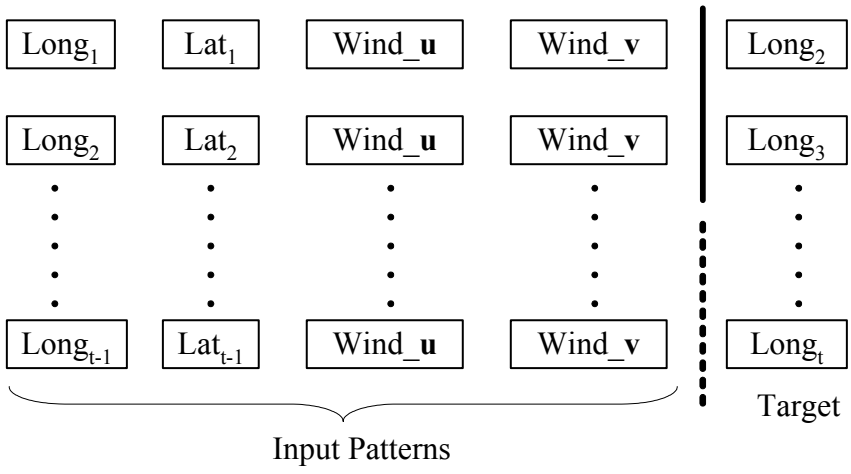
(5) นำข้อมูลลมทั้ง 48 ชุดมาทดสอบกับตำแหน่งต่างๆในขอบเขตพื้นที่ที่กำหนดไว้ 10 ตำแหน่งด้วยแบบจำลองกระแสน้ำพรินซ์ตัน โดยในแต่ละตำแหน่งจะให้เอาต์พุตออกมา 12 ค่าซึ่งหมายถึง ตำแหน่งการเคลื่อนที่ของวัตถุในเวลา 12 ชั่วโมง ตามขนาดของข้อมูลลมที่ป้อนเข้าไปทุกๆ 12 ชั่วโมงเช่นกัน

(6) เก็บรวบรวมข้อมูลที่ได้จากแบบจำลองกระแสน้ำพรินซ์ตัน เพื่อเป็นข้อมูลสำหรับสำหรับสอนและทดสอบแบบจำลองกระแสน้ำที่ได้พัฒนาขึ้น

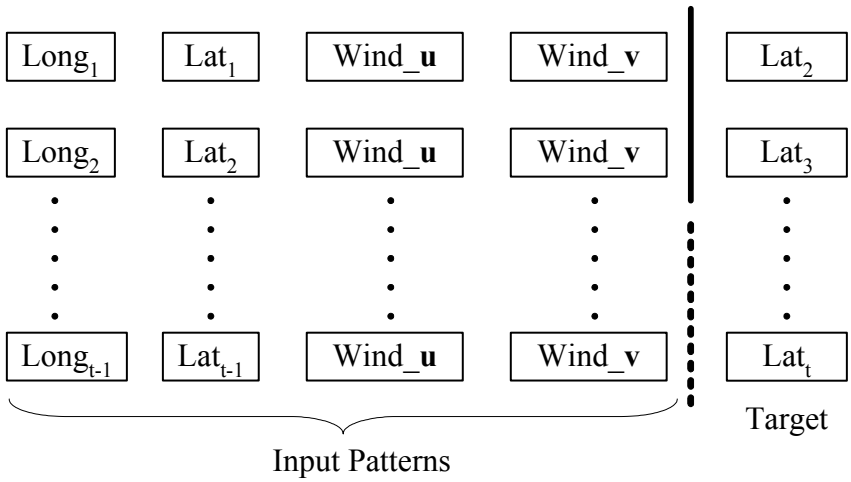
4.3 รายละเอียดขั้นตอนการสอนซัพพอร์ตเวกเตอร์แมทชีน

ในส่วนของรายละเอียดในการออกแบบแบบจำลองกระแสน้ำและขั้นตอนการสอน (Training) ซัพพอร์ตเวกเตอร์แมทชีนนั้น มีลำดับขั้นตอนดำเนินการแบ่งออกเป็นข้อๆ ดังนี้

(1) ซัพพอร์ตเวกเตอร์แมทชีนจะให้เอาต์พุตออกมาได้ครั้งละ 1 ค่าต่อข้อมูลอินพุต 1 ชุด แต่งานวิจัยนี้ต้องการข้อมูลเอาต์พุต 2 ค่า (ละติจูด, ลองจิจูด) ดังนั้นจึงจำเป็นต้องออกแบบแบบจำลองขึ้นมา 2 ชุด เพื่อให้ได้แบบจำลองสำหรับให้ค่าลองจิจูด (Longitude_model) และให้ค่าละติจูด (Latitude_model) ตามลำดับ รูปแบบการจัดเตรียมข้อมูลสำหรับการฝึกสอนแสดงรายละเอียดดังรูป 4.6-4.7



รูป 4.6 รูปแบบข้อมูลสำหรับการฝึกสอนเพื่อสร้าง Longitude_model



รูป 4.7 รูปแบบข้อมูลสำหรับการฝึกสอนเพื่อสร้าง Latitude _model

(2) รูปแบบการจัดเตรียมเซตของข้อมูลอินพุตสำหรับฝึกสอนจากรูป 4.6-4.7 มีการใช้ลักษณะเด่น (Feature) ทั้งหมด 4 ลักษณะเด่นโดยมีรูปแบบดังนี้

$$x1_t = f(x1_{t-1}\ x2_{t-1}\ \ x3\ \ x4) \tag{4.1}$$

$$x2_t = f(x1_{t-1}\ x2_{t-1}\ \ x3\ \ x4) \tag{4.2}$$

โดยที่

$x1$ คือ ข้อมูลตำแหน่งลองจิจูด

$x2$ คือ ข้อมูลตำแหน่งละติจูด

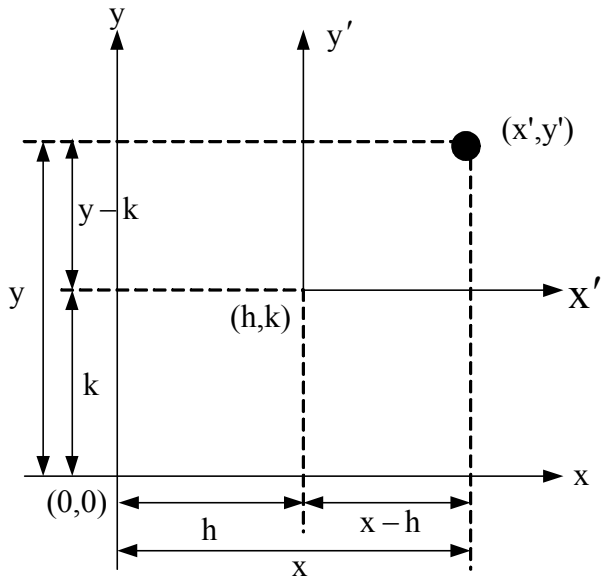
$x3$ คือ ข้อมูลลมในแนวแกน **u**

$x4$ คือ ข้อมูลลมในแนวแกน **v**

สมการ (4.1) หมายถึง ข้อมูล $x1$ ที่เวลาปัจจุบัน (t) เกิดจากข้อมูล $x1$ กับข้อมูล $x2$ ที่เวลาย้อนหลัง 1 ค่า (t–1) และข้อมูล $x3$ กับ $x4$

สมการ (4.2) หมายถึง ข้อมูล $x2$ ที่เวลาปัจจุบัน (t) เกิดจากข้อมูล $x1$ กับข้อมูล $x2$ ที่เวลาย้อนหลัง 1 ค่า (t–1) และข้อมูล $x3$ กับ $x4$

(3) การกำหนดตำแหน่งอ้างอิงของวัตถุบริเวณพื้นที่อ่าวไทย มีการเปลี่ยนแปลงพิกัดของข้อมูลโดยใช้วิธีการย้ายแกนอ้างอิงใหม่ โดยอาศัยเทคนิควิธีการทางเรขาคณิตย้ายแกนจุดเริ่มต้นไปไว้ที่ตำแหน่ง เส้นลองจิจูดที่ 99° ตะวันออก เส้นละติจูดที่ 5° เหนือ ดังรูป 4.8



รูป 4.8 การย้ายแกนเพื่อกำหนดจุดอ้างอิงใหม่

เนื่องจากตำแหน่งของข้อมูลเริ่มต้นจะเริ่มจากตำแหน่งเส้นลองจิจูดที่ 0° เส้นละติจูดที่ 90° ใต้ ส่วนพื้นที่ที่ทำการวิจัยจะเริ่มจากตำแหน่งเส้นลองจิจูดที่ 99° ตะวันออก เส้นละติจูดที่ 5° เหนือ ดังนั้นจึงกำหนดให้ตำแหน่งดังกล่าวเป็นตำแหน่งเริ่มต้น (h,k) เพราะฉะนั้นตำแหน่งที่ใช้สำหรับอ้างอิงใหม่ (x′,y′) ที่เกิดจากการย้ายแกนจะมีสมการเป็น

$$x' = x - h \tag{4.3}$$

$$y' = y - k \tag{4.4}$$

โดยกำหนดให้ $h = 99, k = 5$ และเมื่อแปลงกลับไปเป็นตำแหน่งจริงจะได้

$$x = x' + h \tag{4.5}$$

$$y = y' + k \tag{4.6}$$

(4) รวบรวมข้อมูลการเคลื่อนที่ของวัตถุที่ได้จากแบบจำลองกระแสน้ำพรินซ์ตันในระยะเวลา 12 ชั่วโมง จากจุดพิกัดต่างๆบริเวณอ่าวไทยได้ทั้งหมด 575 รูปแบบ และใช้ขั้นตอนวิธีการครอสแวลิเดชันแบบ 5 กลุ่ม (Five-fold Cross Validation) กับข้อมูลดังกล่าวในการฝึกสอนเพื่อเลือกแบบจำลองที่ดีที่สุด วิธีการดังกล่าวจะแบ่งข้อมูลทั้งหมดออกเป็น 5 ส่วน โดยจัดเป็นข้อมูลสำหรับฝึกสอน 4 ส่วน และข้อมูลสำหรับทดสอบ 1 ส่วน ซึ่งคิดเป็นเปอร์เซ็นต์เท่ากับ 80 และ 20 เปอร์เซ็นต์ ตามลำดับ

(5) นำข้อมูลทั้งหมดมาทำการนอร์แมลไลซ์ (Normalize) ให้มีค่าอยู่ในช่วง [–1 1] ก่อนที่จะนำไปใช้ฝึกสอน ซึ่งเป็นช่วงการทำงานของซัพพอร์ตเวกเตอร์แมทชีน

(6) กำหนดค่าพารามิเตอร์ของการใช้งานของซัพพอร์ตเวกเตอร์แมทชีน ได้แก่ การเลือกใช้ฟังก์ชันเคอร์เนล ค่าพารามิเตอร์ของฟังก์ชันเคอร์เนล ค่าคลาดเคลื่อนที่ยินยอมได้ (ε) และค่าคงที่ (C) ดังนี้

(6.1) เลือกใช้ฟังก์ชันเคอร์เนลแบบฟังชันเรเดียลเบซิส (RBF Kernel Function) เนื่องจากเป็นที่นิยมใช้กับงานประเภทการวิเคราะห์การถดถอย หรืองานด้านการประมาณค่าฟังก์ชัน [8,14,40] ซึ่งมีรูปแบบดังนี้

$$K(\mathbf{x},\mathbf{x}_i) = \exp(-\| \mathbf{x} - \mathbf{x}_i \|^2 / 2\sigma^2) \tag{4.7}$$

โดยที่ $\mathbf{x}$ คือ เวกเตอร์อินพุต

$\mathbf{x}_i$ คือ ซัพพอร์ตเวกเตอร์ที่ i

σ คือ ค่าพารามิเตอร์ที่ใช้กำหนดสมรรถนะในการเรียนรู้ของฟังก์ชันเคอร์เนลแบบฟังก์ชันเรเดียลเบซิส

สิ่งที่ต้องกำหนดค่าก็คือ σ ซึ่งในเบื้องต้นจะกำหนดให้ค่า $\sigma = 0.5$

(6.2) ข้อมูลเอาต์พุตที่ได้จากแบบจำลองกระแสน้ำมีความละเอียดเป็นเลขทศนิยม 4 ตำแหน่งซึ่งมีความหมายดังนี้

1°	=	60	ไมล์ทะเล
1 ไมล์ทะเล	=	1,800	เมตร
1°	=	108,000	เมตร
0.0001°	=	10.80	เมตร

จากรายละเอียดข้างต้นจะได้ว่าความละเอียดของข้อมูลเอาต์พุตจะมีค่าเท่ากับ 10.80 เมตร ดังนั้นค่าพารามิเตอร์อีกตัวหนึ่งของซัพพอร์ตเวกเตอร์แมทชีนสำหรับการถดถอยคือ ε ซึ่งเป็นการกำหนดค่าคลาดเคลื่อนที่ยินยอมได้ ในที่นี้กำหนดให้ค่าเอาต์พุตของแบบจำลองมีความถูกต้องถึงทศนิยมตำแหน่งที่ 4 ดังนั้นจึงกำหนดให้ $\varepsilon = 1.0 \times 10^{-5}$

(6.3) ค่าคงที่ C ($C > 0$) เป็นค่าพารามิเตอร์อีกตัวหนึ่ง สำหรับใช้คุมค่าคลาดเคลื่อน (Regularization) ในการสร้างระนาบเกินที่เหมาะที่สุด (Optimal Hyperplane) เพื่อใช้แทนกลุ่มข้อมูล ในเบื้องต้นนี้จะกำหนดให้ $C = 100,000$

(7) ทำการฝึกสอนและทดสอบแบบจำลองที่ได้โดยใช้ขั้นตอนวิธีการครอสแวลิเดชันแบบ 5 กลุ่ม เพื่อคัดเลือกแบบจำลองที่ดีที่สุดโดยใช้ค่าผิดพลาดกำลังสองเฉลี่ย (Mean Square Error: MSE) เป็นเกณฑ์ในการเปรียบเทียบสมรรถนะของแบบจำลอง ได้ผลดังตาราง 4.2-4.3 ผลการเปรียบเทียบในรูปแบบของกราฟเส้นเชิงประกอบแสดงดังรูป 4.9-4.10

ตาราง 4.2 ผลการทำครอสแวลิเดชันแบบ 5 กลุ่มเพื่อเลือกแบบจำลองที่ให้ค่าลองจิจูด

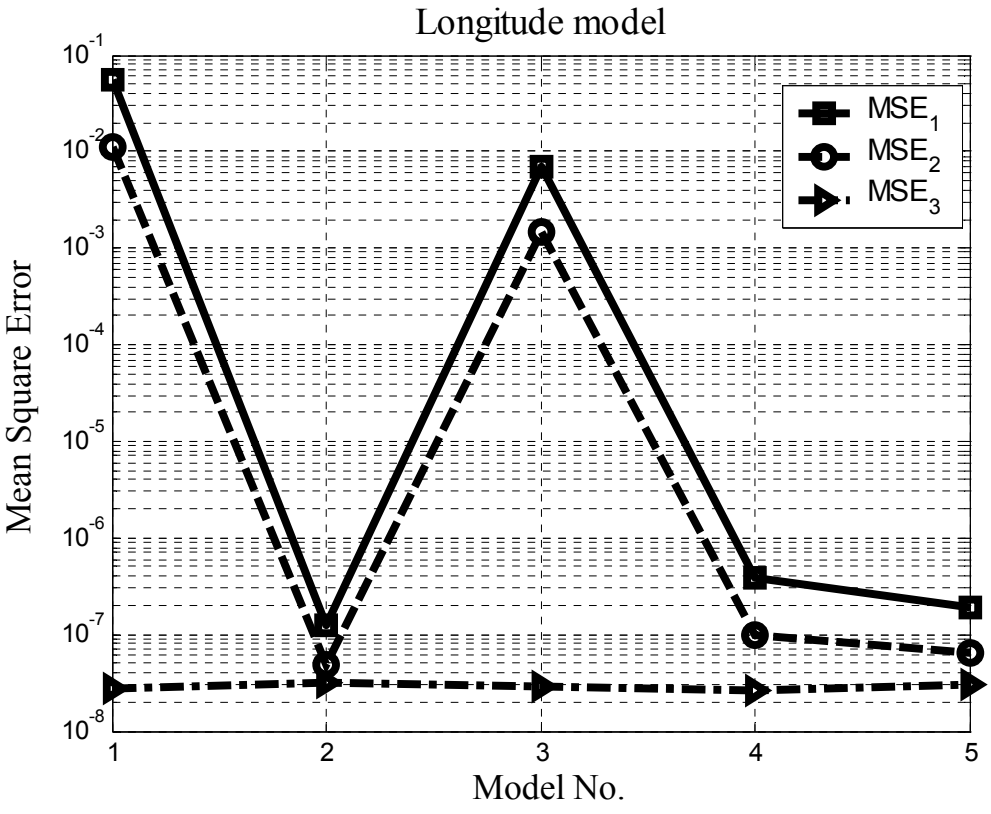
($\sigma = 0.5$, $C = 100{,}000$, $\varepsilon = 1.0\times10^{-5}$)

Longitude_model			
No.	**MSE_1**	**MSE_2**	**MSE_3**
1.	5.6534e-002	1.1365e-002	2.7194e-008
2.	1.2401e-007	4.9615e-008	3.1016e-008
3.	6.8354e-003	1.4875e-003	2.9011e-008
4.	3.8245e-007	9.7913e-008	2.6779e-008
5.	1.9434e-007	6.3032e-008	3.0206e-008

หมายเหตุ MSE_1 ทดสอบเฉพาะ 1 โฟลด์ที่ได้เตรียมไว้ (20%)

MSE_2 ทดสอบทั้ง 5 โฟลด์ (100%)

MSE_3 ทดสอบ 4 โฟลด์ที่นำมาฝึกสอน (80%)

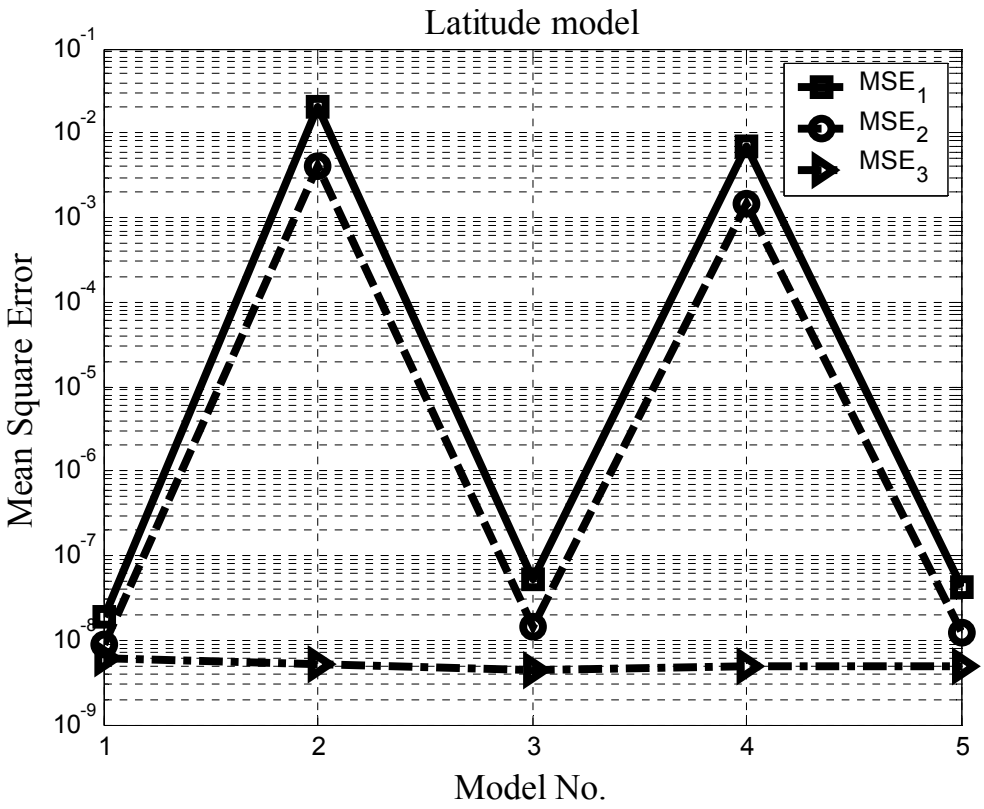


รูป 4.9 ผลการทำครอสแวลิเดชันแบบ 5 กลุ่มเพื่อคัดเลือก Longitude_model

ตาราง 4.3 ผลการทำครอสแวลิเดชันแบบ 5 กลุ่มเพื่อเลือกแบบจำลองที่ให้ค่าละติจูด

(σ= 0.5, C = 100,000, ε= 1.0×10^{-5})

Latitude_model			
No.	**MSE_1**	**MSE_2**	**MSE_3**
1.	1.9251e-008	8.7030e-009	6.0660e-009
2.	2.0041e-002	4.0165e-003	5.3084e-009
3.	5.3367e-008	1.4136e-008	4.3282e-009
4.	7.0452e-003	1.4564e-003	4.8113e-009
5.	4.3015e-008	1.2538e-008	4.9189e-009



รูป 4.10 ผลการทำครอสแวลิเดชันแบบ 5 กลุ่มเพื่อคัดเลือก Latitude_model

สรุปผลการคัดเลือกแบบจำลองที่เหมาะที่สุด ด้วยขั้นตอนวิธีการครอสแวลิเดชันแบบ 5 กลุ่ม อาศัยผลการทดลองจากตาราง 4.2-4.3 โดยพิจารณาค่า MSE_1 เป็นหลัก ซึ่งเป็นค่าผิดพลาดกำลังสองเฉลี่ย ของแบบจำลองที่ได้จากการทดสอบด้วยข้อมูลที่ไม่ได้นำมาฝึกสอน จึงเลือกแบบจำลองที่ให้ค่าลองจิจูดหมายเลข 2 (Longitude_model No.2) และเลือกแบบจำลองที่ให้ค่าละติจูดหมายเลข 1 (Latitude_model No.1) ซึ่งเป็นแบบจำลองที่ให้ค่า MSE_1 ต่ำที่สุด

(8) นำแบบจำลองที่คัดเลือกได้ในเบื้องต้นมาทำการทดลองหาค่า σ และค่า C ที่เหมาะที่สุดต่อไป โดยการทดลองเปลี่ยนค่า σ และค่า C ไปเรื่อยๆ จนได้ค่าที่เหมาะที่สุดที่สามารถนำไปใช้งานได้ ซึ่งได้ผลตามการทดลองดังนี้

(8.1) ทดลองเปลี่ยนค่า σ ของแบบจำลอง 10 ค่า คือ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 และ 1.0 ผลจากการทดลองแสดงในตาราง 4.4-4.5 รูป 4.11-4.12

(8.2) ทดลองเปลี่ยนค่า C ของแบบจำลอง 10 ค่าคือ 1,000; 5,000; 10,000; 50,000; 100,000; 200,000; 300,000; 400,000; 500,000 และ 1,000,000 ผลการทดลองดังแสดงในตาราง 4.6-4.7 รูป 4.13-4.14

ตาราง 4.4 ผลการทดลองการเปลี่ยนค่า σ เพื่อคัดเลือกแบบจำลองค่าลองจิจูด

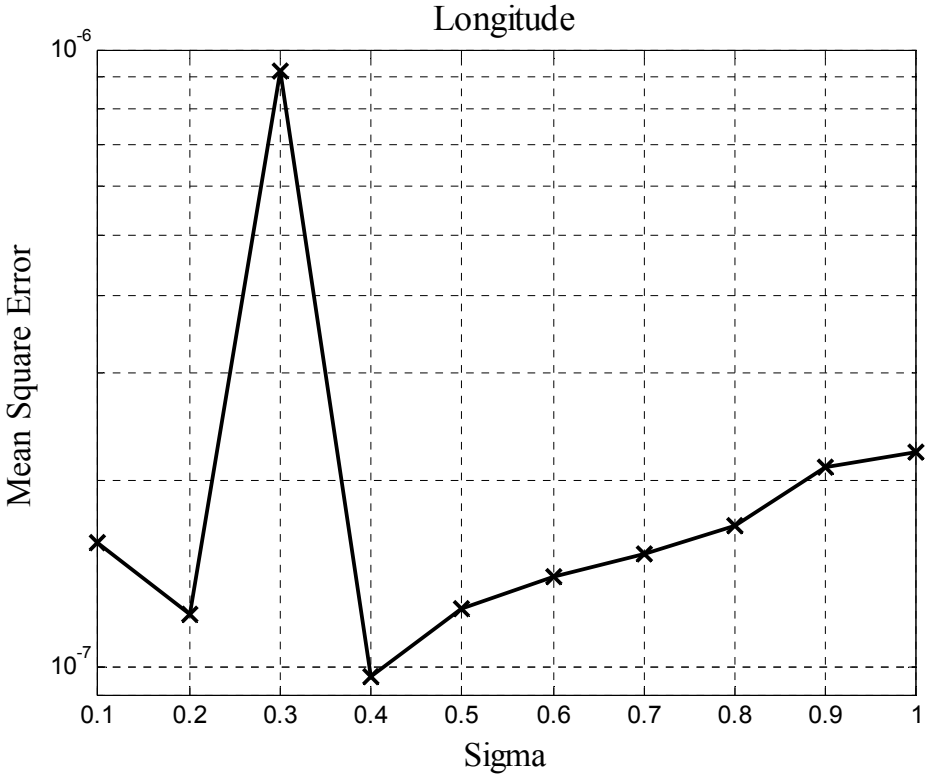
(C = 100,000, $\varepsilon = 1\times10^{-5}$)

σ	**Longitude_model (MSE.)**
0.1	1.587e-007
0.2	1.214e-007
0.3	9.231e-007
0.4	9.625e-008
0.5	1.240e-007
0.6	1.398e-007
0.7	1.517e-007
0.8	1.687e-007
0.9	2.107e-007
1.0	2.227e-007

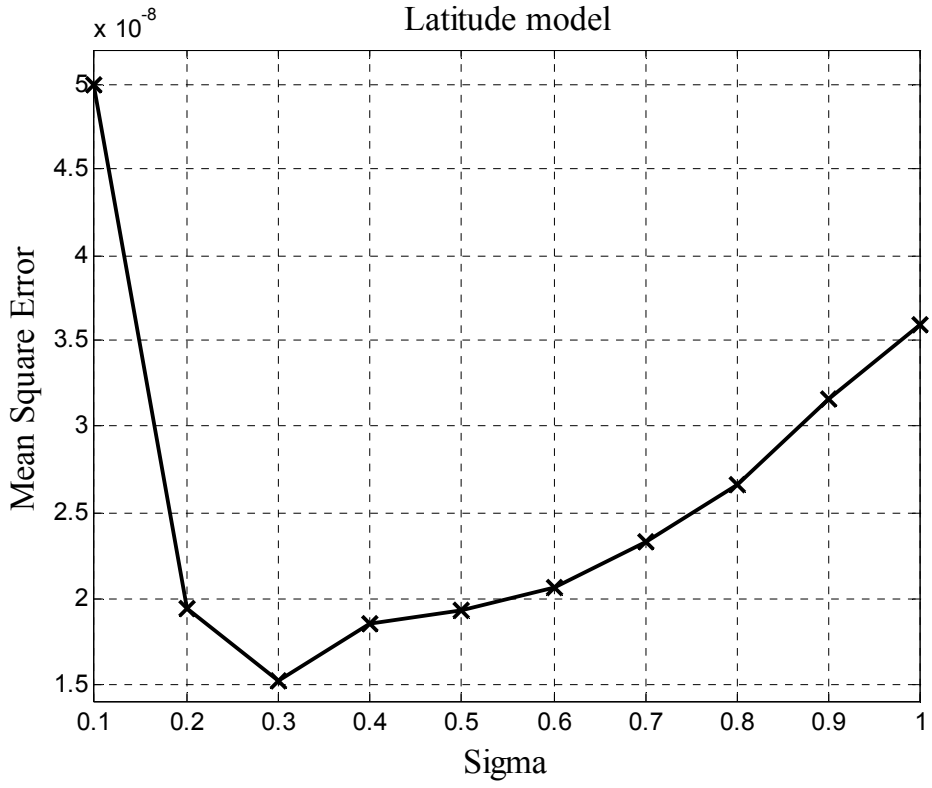
ตาราง 4.5 ผลการทดลองการเปลี่ยนค่า σ เพื่อคัดเลือกแบบจำลองค่าละติจูด

(C = 100,000, $\varepsilon = 1\times10^{-5}$)

σ	**Latitude_model (MSE.)**
0.1	4.9889e-008
0.2	1.9355e-008
0.3	1.5181e-008
0.4	1.8452e-008
0.5	1.9251e-008
0.6	2.0637e-008
0.7	2.3226e-008
0.8	2.6618e-008
0.9	3.1581e-008
1.0	3.5999e-008



รูป 4.11 ผลการทดลองการเปลี่ยนค่า σ เพื่อคัดเลือกแบบจำลองค่าลองจิจูด

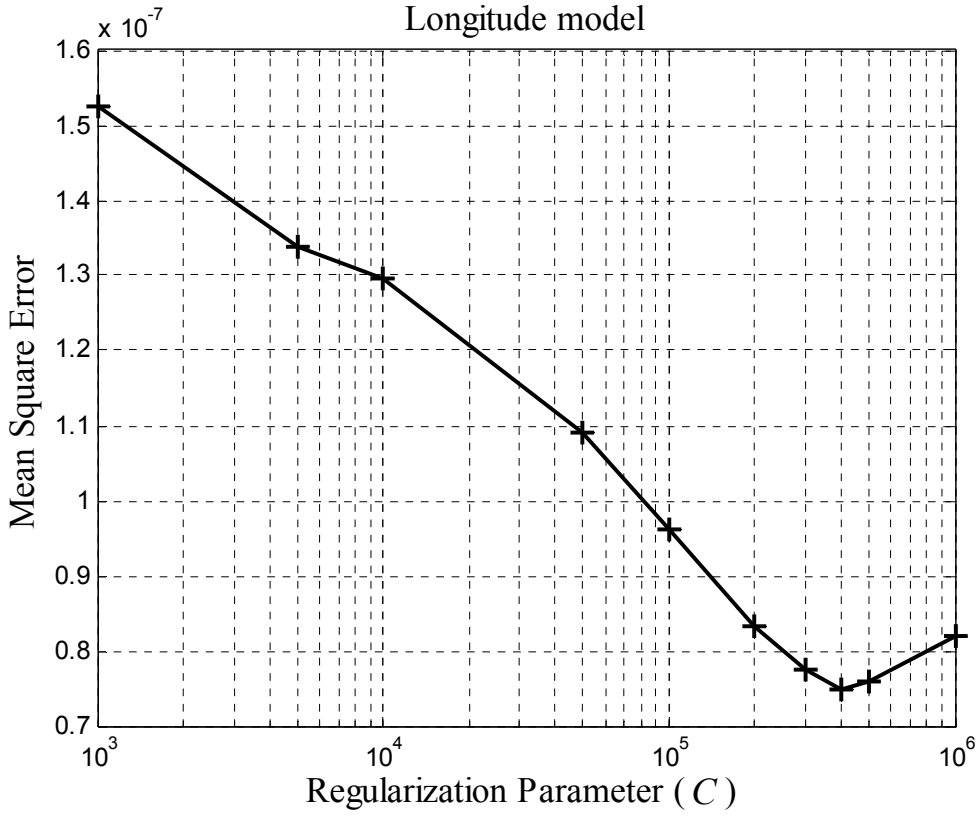


รูป 4.12 ผลการทดลองการเปลี่ยนค่า σ เพื่อคัดเลือกแบบจำลองค่าละติจูด

สรุปผลการคัดเลือกค่า σ ที่เหมาะที่สุด อาศัยผลการทดลองโดยพิจารณาค่า MSE จากรูป 4.11-4.12 พบว่าค่า σ ที่ทำให้ค่า MSE ของแบบจำลองค่าลองจิจูดต่ำที่สุดเท่ากับ 0.4 และของแบบจำลองค่าละติจูดต่ำที่สุดเท่ากับ 0.3 ตามลำดับ

ตาราง 4.6 ผลการทดลองการเปลี่ยนค่า C เพื่อคัดเลือกแบบจำลองค่าลองจิจูด ($\varepsilon = 1\times10^{-5}$)

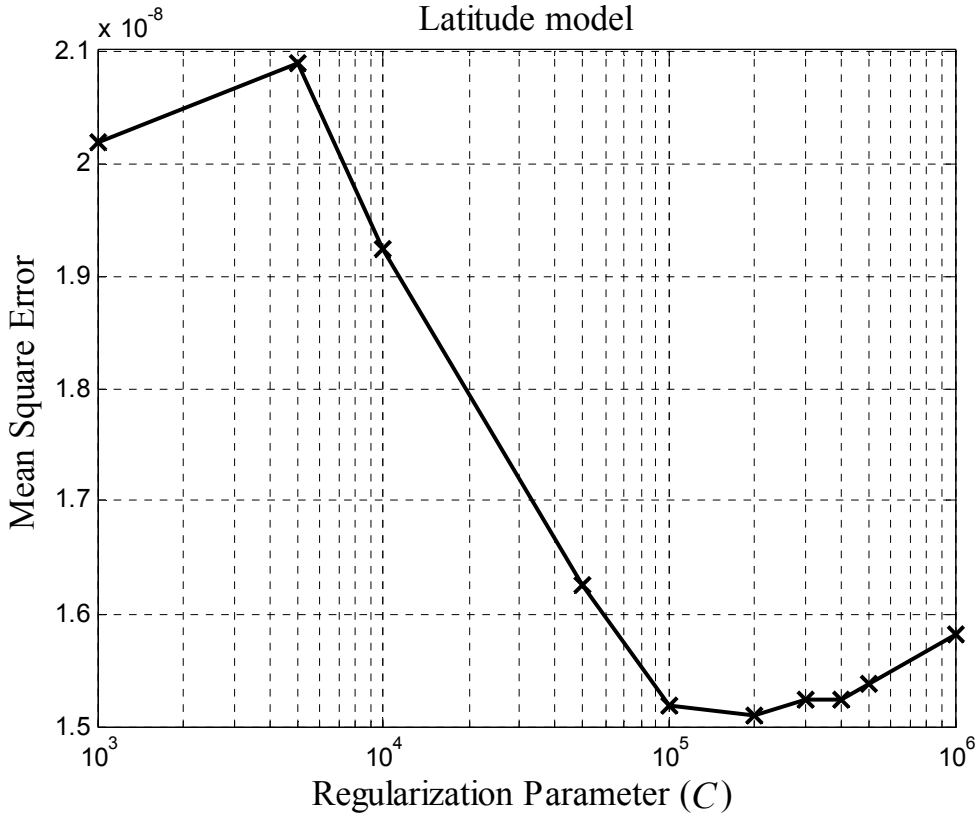
	Longitude_model
C	**MSE.** ($\sigma = 0.4$)
1,000	1.5248e-007
5,000	1.3390e-007
10,000	1.2953e-007
50,000	1.0918e-007
100,000	9.6247e-008
200,000	8.3368e-008
300,000	7.7421e-008
400,000	7.4996e-008
500,000	7.5936e-008
1,000,000	8.1965e-008



รูป 4.13 ผลการทดลองการเปลี่ยนค่า C เพื่อคัดเลือกแบบจำลองค่าลองจิจูด

ตาราง 4.7 ผลการทดลองการเปลี่ยนค่า C เพื่อคัดเลือกแบบจำลองค่าละติจูด (ε=1×10^{-5})

C	**Latitude_model**
	MSE. (σ = 0.3)
1,000	2.0180e-008
5,000	2.0885e-008
10,000	1.9238e-008
50,000	1.6246e-008
100,000	1.5181e-008
200,000	1.5091e-008
300,000	1.5239e-008
400,000	1.5231e-008
500,000	1.5369e-008
1,000,000	1.5808e-008



รูป 4.14 ผลการทดลองการเปลี่ยนค่า C เพื่อคัดเลือกแบบจำลองค่าละติจูด

สรุปผลการคัดเลือกค่า C ที่เหมาะที่สุด อาศัยผลการทดลองโดยพิจารณาค่า MSE จากรูป 4.13-4.14 พบว่าค่า C ที่ทำให้ค่า MSE ของแบบจำลองค่าลองจิจูดต่ำที่สุดเท่ากับ 400,000 และของแบบจำลองค่าละติจูดต่ำที่สุดเท่ากับ 200,000 ตามลำดับ

จากผลการทดลองออกแบบและคัดเลือกแบบจำลองที่เหมาะที่สุด โดยใช้ขั้นตอนวิธีการครอสแวลิเดชันแบบ 5 กลุ่ม และทำการทดลองหาค่าพารามิเตอร์ที่เหมาะสม ให้กับแบบจำลองที่เลือกได้โดยการใช้ค่า MSE เป็นเกณฑ์ตัดสินวัดความสามารถในการเรียนรู้ของแบบจำลอง ได้ผลสรุปออกมาดังนี้

ตาราง 4.8 ผลสรุปการทดลองเพื่อคัดเลือกแบบจำลองกระแสน้ำที่เหมาะที่สุด

Longitude_model (No.2)			**Latitude_model (No.1)**		
σ	***C***	**MSE.**	σ	***C***	**MSE.**
0.4	400,000	7.4996e-008	0.3	200,000	1.5091e-008

ในบทต่อไปจะนำแบบจำลองที่คัดเลือกได้ ไปทำการทดสอบความถูกต้องเปรียบเทียบผลกับแบบจำลองกระแสน้ำพรินซ์ตัน โดยใช้ข้อมูลที่ได้เตรียมไว้สำหรับการทดสอบและข้อมูลที่มีการเก็บบันทึกไว้จากการทดสอบ และใช้งานจริงในทะเล