บทที่ 4

การพัฒนาแบบจำลองการลดมาตราส่วน (SDSM) ของฝนรายวัน

4.1 แบบจำลองการลดมาตราส่วน (SDSM)

วิธีการลดมาตราส่วน (downscaling) จะต้องนำมาใช้เมื่อการจำลองตัวแปรของระบบ GCM (หรือ RCM) ซึ่งใช้สำหรับจำลองผลกระทบที่เกิดขึ้นนั้น ไม่เป็นไปตามสภาพที่ความเป็นจริงตามมาตราส่วนที่สนใจศึกษา อาจเกิดจากมาตราส่วนที่ต่ำกว่าความละเอียดที่ได้จากแบบจำลองสภาพที่ภูมิอากาศ หรือ อาจเกิดจากความไม่เพียงพอของตัวแบบจำลอง การลดมาตราส่วนนี้อาจใช้การสร้างภาพที่ฉายต่างๆ ที่ไม่สามารถทำได้จาก GCMs และ RCMs ในการจำลอง อย่างไรก็ตาม GCM สามารถอธิบายตัวแปรในมาตราส่วนขนาดใหญ่ซึ่งมีความสำคัญอย่างมากสำหรับกระบวนการในระดับภูมิภาค ในทางปฏิบัติการเลือกวิธีการลดมาตราส่วนจะขึ้นอยู่กับข้อมูลGCMที่มีและที่สามารถหามาได้จากการเก็บข้อมูล ซึ่งล้วนจำเป็นต่อการสร้างภาพที่ฉายต่าง ๆ ของสภาพที่ภูมิอากาศในอนาคต

ซอฟแวร์SDSMได้ลดขั้นตอนในการลดมาตราส่วนแบบสถิติของสภาพที่อากาศรายวันให้เหลือเพียง 7 ขั้นตอน (กล่องที่เส้นทึบ ในภาพที่ 4. 1) คือ

(1) การควบคุมคุณภาพที่และการแปรสภาพที่ข้อมูล

(2) การคัดกรองตัวแปรของการทำนายการลดมาตราส่วน

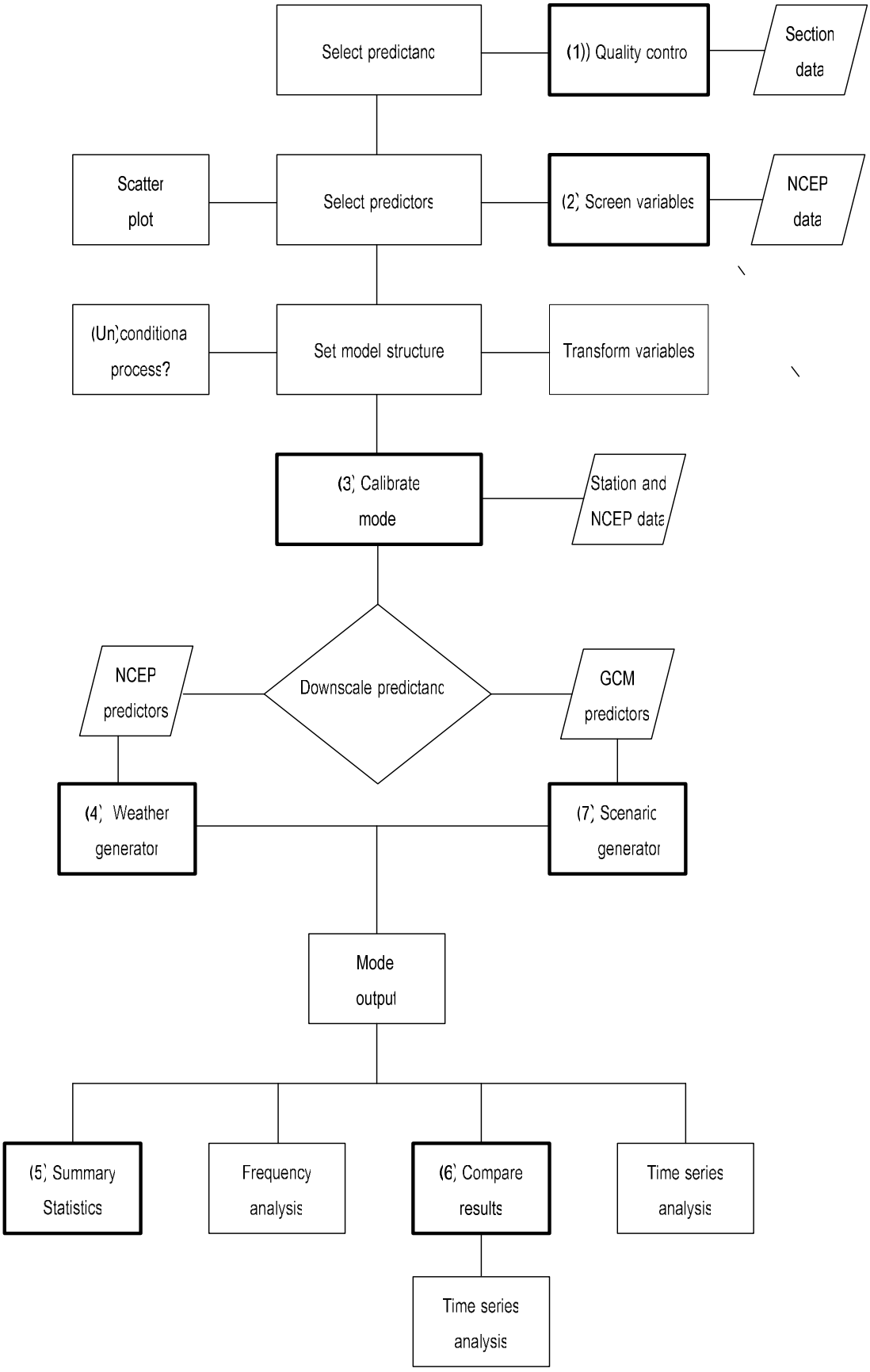
(3) คำนวณแบบจำลองปรับเทียบ

(4) การสังเคราะห์ข้อมูล

(5) วิเคราะห์ทางสถิติ

(6) แสดงผลลัพธ์จากแบบจำลองเป็นกราฟฟิก

(7) การสร้างภาพที่ฉาย



ภาพที่ 4.1 การทำงานของแบบจำลองการลดมาตราส่วน (SDSM) เวอร์ชั่น4.1

เทคนิคลดมาตราส่วน SDSM กล่าวได้ว่าเป็นลูกผสมระหว่างการรวมกันของสภาพที่อากาศแบบสถิติ (stochastic) และ วิธีการถดถอย (regression) เนื่องจากรูปแบบการหมุนเวียนสภาพที่อากาศในสเกลใหญ่และค่าตัวแปรของความชื้นในบรรยากาศเป็นสัดส่วนโดยตรงกับค่าพารามิเตอร์ของอัตราส่วนสเกลระดับท้องถิ่น ปัจจุบันวิธีการลดมาตราส่วนตามกลไกลของ SDSM ถูกนำไปใช้ในงานอุตุนิยมวิทยา ชลประทาน ด้านสิ่งแวดล้อม และในงานด้านภูมิศาสตร์ทั้งในทวีปยุโรป อเมริกาเหนือ และเอเชียตะวันออกเฉียงใต้

ขั้นตอนการทำงานของแบบจำลอง SDSM ดังที่กล่าวไว้แล้วก่อนหน้านี้แล้วว่าการลดมาตราส่วนแบบสถิติของสภาพที่อากาศรายวัน ให้เหลือเพียง 7 ขั้นตอน ซึ่งจุดประสงค์การทำงานทั้ง 7 ขั้นตอนมีดังนี้

(1) การควบคุมคุณภาพที่ข้อมูล

สถานีวัดน้ำฝนจำนวนน้อยมากที่มีข้อมูลเสร็จสิ้นสมบูรณ์ 100% หรือมีข้อมูลถูกต้องหมด การจัดการกับข้อมูลที่หายไปและที่บกพร่องมีความจำเป็นในทางปฏิบัติอย่างมาก การควบคุมคุณภาพที่ของข้อมูลทำได้โดยการ การนำค่าข้อมูลฝนที่ทำการเติมข้อมูลเรียบร้อยแล้ว ให้แบบจำลองSDSM ทำการตรวจสอบ โดยแบบจำลองจะทำการวิเคราะห์ ค่าต่ำสุด ค่าสูงสุด ค่าเฉลี่ย จำนวนวันรวม26ปี (ค.ศ. 1976-ค.ศ. 2001) อีกทั้งตรวจสอบว่ามีการขาดหายข้องข้อมูลบ้างหรือไม่

(2) การคัดกรองตัวแปรNCEP สำหรับการทำนายการลดมาตราส่วน

จุดประสงค์หลักของการทำงานการคัดกรองตัวแปรคือจะช่วยให้เลือกความเหมาะสมของตัวแปรที่ทำนายต่อการลดมาตราส่วนในการคัดกรองจะพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์แยกส่วน (partial correlation) ซึ่งเป็นการหาความสำคัญระหว่างตัวแปร NCEP ตัวนั้นกับฝนรายของสถานีที่พิจารณา ซึ่งถูกควบคุมโดยตัวแปร NCEP ตัวอื่นๆ ที่แบบจำลองSDSM คำนวณได้ซึ่งในแบบจำลองทำเพียงครั้งเดียว

(3) แบบจำลองปรับเทียบ

แบบจำลองจะทำการคำนวณตัวแปรของสมการการถดถอยหลายครั้งเพื่อให้ได้ผลดีที่สุดระหว่างสถานีที่พิจารณาและตัวแปร NCEP ที่ผ่านการคัดกรองแล้ว ซึ่งแบบจำลองได้ทำการลดมาตราส่วนจากระดับโลกมาเป็นระดับท้องถิ่นแล้ว และจะเก็บข้อมูลไว้ เพื่อใช้ในการสังเคราะห์ต่อไป

(4) การสังเคราะห์ข้อมูล

การทำงานของการสังเคราะห์อากาศจะสร้างทั้งชุดของอากาศประจำวันสังเคราะห์ซึ่งจะเป็นเป็นอนุกรมเฝ้าสังเกต (หรือ NCEP re–analysis) ตัวแปรการทำนายเกี่ยวกับบรรยากาศโดยใช้ได้ขบวนการการพิสูจน์ความจริงของแบบจำลองปรับเทียบและการสังเคราะห์ของเวลาสมมุติซึ่งเป็น

อนุกรมสำหรับแสดงสภาวะอากาศตามฤดูกาลโดยแบบจำลองปรับเทียบ และ SDSM จะเชื่อมโดยการทำนายทั้งหมดโดยอัตโนมัติ การสังเคราะห์ฝนจะแบ่งการสังเคราะห์ ออกเป็น2ช่วงเวลาด้วยกัน คือ ปี ค.ศ. 1976-1990 (15ปี) สำหรับการปรับเทียบและจากปี ค.ศ. 1991-2001 (11ปี) สำหรับสอบความสมเหตุสมผลของแบบจำลอง SDSM ซึ่งแบบจำลองมีข้อจำกัดในการสังเคราะห์ข้อมูลได้สูงสุด 100 ชุดข้อมูล ในการสังเคราะห์นี้จะสั่งการให้แบบจำลองทำการสังเคราะห์ 100 ครั้ง เพื่อให้ได้ค่าที่ละเอียดต่อการนำ ไปทำการวิเคราะห์ข้อมูลต่อไป

(5) การวิเคราะห์ข้อมูล

SDSM จัดเตรียมวิธีการลดมาตราส่วนภาพที่ฉาย และข้อมูลอากาศตามฤดูกาลที่เฝ้าสังเกตกับข้อสรุปทางสถิติและคัดกรองการวิเคราะห์ความถี่ สำหรับผลลัพธ์แบบจำลองจะต้องถูกเจาะจงในทางกลับกันในการวิเคราะห์ครั้งหนึ่ง SDSM จะแสดงรวมทุกอย่าง/ทุกเดือน/ตามฤดูการ/ค่าเฉลี่ยประจำปี ซึ่งในการวิเคราะห์ข้อมูลของฝนได้ใช้ค่าทางสถิติดังนี้ คือ ค่าเฉลี่ย (mean), ค่าความแปรปรวน (variance), ค่าความเบ้ (skewness), เปอร์เซ็นต์ฝนตก(percentage wet), ค่าเฉลี่ยของช่วงเวลาที่ฝนไม่ตก (mean dry spell length), ค่าเฉลี่ยของช่วงเวลาที่ฝนตก (mean wet spell length), จำนวนครั้งสัมพัทธ์ที่ฝนตก (mean wet-day persistence), จำนวนครั้งสัมพัทธ์ที่ฝนไม่ตก (mean dry-day persistence)

(6) การวิเคราะห์กราฟฟิค

ภายใต้เงื่อนไข SDSM จะแสดงทางเลือกสำหรับการวิเคราะห์กราฟฟิค 3 ทางเลือกคือ การวิเคราะห์ความถี่, การเปรียบเทียบผลลัพธ์ และการแสดงการวิเคราะห์ซึ่งเป็นอนุกรมเวลา การแสดงการวิเคราะห์ความถี่โดยจะกำหนดขอบเขตของค่าทางสถิติจากแฟ้มข้อมูลที่เลือก ส่วนผลลัพธ์ของการเปรียบเทียบใช้เพื่อแสดงแผนที่ทางสถิติทุกเดือนที่สังเคราะห์ได้ ซึงต้องมีแฟ้มข้อมูลที่นำเข้า ส่วนการแสดงการวิเคราะห์ซึ่งเป็นอนุกรมเวลาจะผลิตแผนเวลาซึ่งเป็นอนุกรมสำหรับสูงถึงสูงที่สุดของห้าตัวแปรได้แก่ ข้อมูลที่สามารถถูกวิเคราะห์ทุกเดือน,ฤดูกาล,อัตราฤดูหนาวต่อฤดูร้อน,อนุกรมความต่อเนื่อง,และเปอร์เซ็นต์ไทล์ แต่แบบจำลองมีข้อจำกัดในการเปรียบเทียบผลลัพธ์ ได้ทีละคู่

(7) การสร้างภาพที่ฉาย

การทำงานของเครื่องกำเนิดภาพที่ฉายของการทำนายตัวแปรอากาศประจำวันสังเคราะห์ เป็นอนุกรมที่สนับสนุนโดยแบบจำลองอากาศตามฤดูกาล (ที่อันใดอันหนึ่งสำหรับแสดงหรือทดลองอากาศตามฤดูกาลอนาคต) ค่อนข้างดีกว่าการทำนายที่เฝ้าสังเกต การทำงานนี้จะให้ภาพที่ฉายของการทำงานของเครื่องกำเนิดอากาศ สำหรับตัวแปรการทำนาย

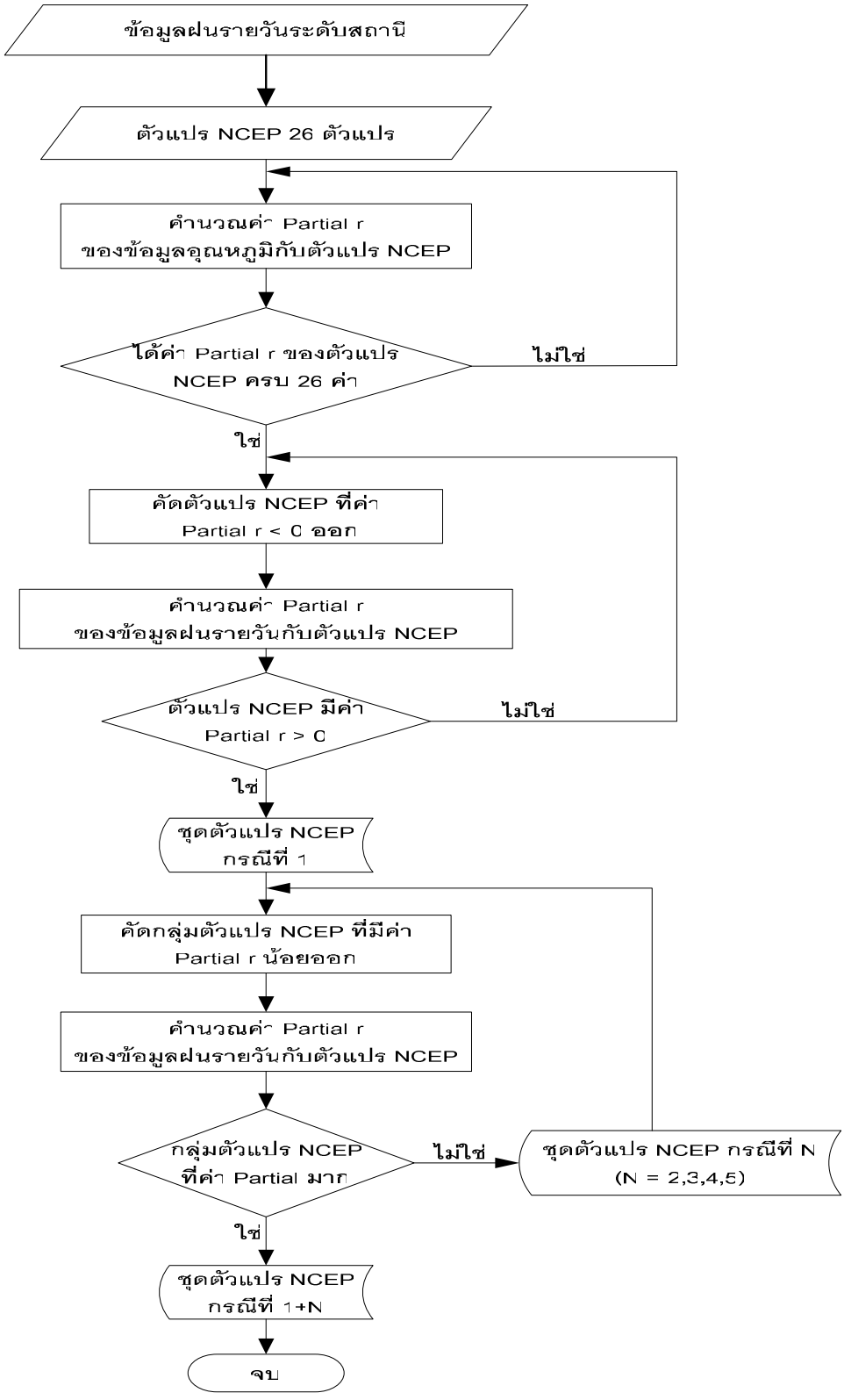
ขั้นตอนทั้ง 7 ขั้นตอนของแบบจำลอง SDSM ข้างต้นนี้สามารถใช้ได้ดีแต่ขั้นตอนที่ (2) ได้ทำการปรับปรุงเนื่องจากเป็นวิธีซึ่งไม่ประหยัดตัวแปร NCEP การศึกษานี้จึงได้เสนอวิธีการคัดกรอง NCEP เพื่อให้ประหยัดจำนวนตัวแปร ตามวิธีที่ได้อธิบายไว้ในหัวข้อ 4.2

4.2 วิธีการคัดกรองข้อมูล NCEP ที่นำเสนอ

หลังจากทำการเติมข้อมูลผลที่ขาดหายไปแล้ว เราก็ได้ทำการควบคุมคุณภาพที่ข้อมูลอีกครั้ง โดยการตรวจสอบว่าสถานีที่นำมาพิจารณามีข้อมูลฝนรายวันครบ 9497 วัน (ปี ค.ศ. 1976-2001) หรือไม่จากการตรวจสอบพบว่าทุกสถานีมีข้อมูลฝนรายวันครบทั้ง 9497 วัน เมื่อมีข้อมูลครบแล้วก็ทำการพัฒนาแบบจำลอง SDSM ของฝนรายวันต่อไป

จุดประสงค์หลักของการคัดกรองตัวแปร คือ เพื่อช่วยในการเลือกชุดตัวแปร NCEP ที่มีความเหมาะสมกับฝนรายวันของสถานีที่ทำการพิจารณา ในการคัดกรองจะพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วน (partial correlation) ซึ่งเป็นการหาความสัมพันธ์ระหว่างตัวแปร NCEP ตัวนั้นกับฝนรายวันรายของสถานีที่พิจารณาโดยถูกควบคุมด้วยตัวแปร NCEP ตัวอื่นๆ

จากภาพที่ 4.2 แสดงถึงขั้นตอนของการคัดกรองชุดตัวแปร NCEP ด้วยค่าสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วนกล่าวคือหลังจากทำการเลือกสถานีเพื่อพิจารณาแล้ว ทำการเลือกชุดตัวแปร NCEP ตามลำดับ (ตารางที่ 4.1) ครั้งละ 12ตัว จนครบทั้ง 26ตัว เพื่อหาค่าสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วนหลังจากแบบจำลองSDSMได้ทำการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วนครบทั้ง 26 ตัว ต่อจะไปนำชุดตัวแปร NCEP ที่มีค่าสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วนที่เป็นมีค่าเป็นบวกมาพิจารณาทั้งหมดซึ่งเมื่อนำตัวแปรNCEP ที่มีค่าสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วน ที่มีค่าเป็นบวกมาพิจารณาอีกครั้งกับตัวแปรNCEP ตัวอื่นที่เป็นบวกเหมือนกัน ก็อาจได้ค่าสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วนที่มีค่าเป็นลบได้จึงต้องทำการคัดกรองตัวแปรใหม่และทำการคัดตัวแปรที่มีค่าสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วนติดลบออก แล้วใส่ตัวแปร NCEP ตัวอื่นเข้าไปแทนทำจนได้ค่าสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วน ของชุดตัวแปร NCEP ที่มีค่าเป็นบวกทั้งหมดและให้ชุดตัวแปร NCEP ชุดนี้ เป็นกรณีที่1 เมื่อได้ชุดตัวแปร NCEP ชุดแรกที่มีค่าสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วนที่มีค่าเป็นบวกทั้งหมดแล้ว ก็ทำการคัดชุดตัวแปรNCEP ที่มีค่าสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วนกลุ่มน้อยออกแล้วก็ให้ชุดตัวแปร NCEP ชุดนั้นเป็นกรณีที่2 ทำการคัดชุดตัวแปร NCEP เช่นเดิมไปจนเหลือชุดตัวแปร NCEP ที่มีค่าสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วน ที่อยู่ในกลุ่มมากที่สุด และให้ชุดตัวแปร NCEP ชุดนั้นเป็นกรณีสุดท้าย



ภาพที่ 4.2 ขั้นตอนการคัดกรองตัว NCEPที่นำเสนอด้วยค่าสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วน

จากการเลือกชุดตัวแปร NCEP ที่ได้มาในแต่ละกรณี ทำได้โดยการเปรียบเทียบดูว่ากรณีใด มีความสำคัญกับฝนรายวันสถานีที่พิจารณามากที่สุด การเปรียบเทียบ จะทำการหาชุดตัวแปร NCEP ที่มีนัยสำคัญต่อฝนรายวันสถานีวัดน้ำฝนรายวันที่ใช้ในการศึกษามากที่สุดโดยจะแบ่งการเปรียบเทียบออกเป็น 2 ช่วงเวลาด้วยกัน คือระหว่างปี ค.ศ. 1976-1990 (15ปี) สำหรับการปรับเทียบ (calibrate) และระหว่างปี ค.ศ. 1991-2001 (11ปี) เพื่อทำการสอบความสมเหตุสมผล (validate) โดยการสอบความสมเหตุสมผลจะถูกนำมาใช้เมื่อไม่สามารถเลือกชุดตัวแปรNCEPกรณีใดได้ในช่วงข้อมูลปรับเทียบ

การปรับเทียบและสอบความสมเหตุสมผลของแบบจำลอง SDSM นี้จะทำการเปรียบเทียบระหว่างคุณลักษณะทางสถิติของฝนรายวัน [ค่าเฉลี่ย (mean), ค่าความแปรปรวน (variance), ค่าความเบ้ (skewness)] ที่คำนวณได้กับข้อมูลฝนอดีต เกณฑ์ที่ใช้ในการหาชุดตัวแปร NCEP ที่เหมาะสมกับฝนรายวันของสถานีต่างๆ มี 3 เกณฑ์ด้วยกัน ได้แก่เกณฑ์ขอบเขตหนึ่งช่วงของความเบี่ยงเบนมาตรฐาน (one-standard deviation bound), เกณฑ์อคติ (bias criterion, β) และ เกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (root mean square error, RMSE) ซึ่งมีนิยามของแต่ละเกณฑ์คือ

เกณฑ์ขอบเขตหนึ่งช่วงของความเบี่ยงเบนมาตรฐาน เป็นเกณฑ์ที่แสดงตำแหน่งของการกระจายของข้อมูลที่พิจารณาว่าคุณลักษณะต่างๆ ของฝนรายวันที่ได้คำนวณได้สามารถเลียนแบบข้อมูลฝนรายวันอดีตได้ดีหรือไม่ กล่าวคือหากว่าคุณลักษณะต่างๆ ของฝนรายวันที่แบบจำลอง SDSM คำนวณได้ อยู่ในขอบเขตหนึ่งช่วงของความเบี่ยงเบนมาตรฐาน, $\mu \pm \sigma$ (เมื่อ μ และσ คือค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคุณลักษณะทางสถิติของฝนรายวันที่คำนวณ $\bar{y}$ ของ y) ก็หมายความว่า แบบจำลอง SDSM สามารถเลียนแบบคุณลักษณะทางสถิติของฝนรายวันอดีต, y ได้ดี โดยจะทำการเปรียบเทียบเป็นเดือนๆ หากเดือนใดค่าเฉลี่ยของคุณลักษณะทางสถิติของฝนรายวันที่คำนวณได้อยู่ในขอบเขตหนึ่งช่วงของความเบี่ยงเบนมาตรฐานก็ให้นับเดือนนั้นเป็น 1คะแนน

เกณฑ์อคติเป็นการวัดความความเบี่ยงเบนของคุณลักษณะทางสถิติของฝนรายวันที่คำนวณได้จากแบบจำลอง SDSM กับข้อมูลฝนรายวันอดีต ดังสมาการ 4.1

$$\beta = \left[E\bar{y}_{SS} \right] - y \qquad (4.1)$$

หากค่าอคติมีค่าน้อยแสดงว่าคุณลักษณะทางสถิติของฝนรายวันที่คำนวณได้สามารถเลียนแบบคุณลักษณะทางสถิติของฝนรายวันอดีตได้ดีในทางกลับกันหากค่าอคติมีค่ามากแสดงว่า

คุณลักษณะทางสถิติของฝนรายวันที่คำนวณเลียนแบบคุณลักษณะทางสถิติของฝนรายวันอดีตได้ไม่ดีพอ

เมื่อคำนวณค่าอคติแล้วก็จะทำการจัดลำดับและให้คะแนน แต่ละกรณีว่ากรณีใดเป็นกรณีที่ดีที่สุด โดยเทียบเป็นแต่ละเดือนของแต่ละกรณี กล่าวคือหากในเดือนนั้นในกรณีใดที่มีค่าอคติค่าห่างจากค่าศูนย์มากที่สุดเป็นกรณีที่แย่ที่สุด โดยให้คะแนนเป็น 1 คะแนน และค่าอคติรองลงมาก็ให้คะแนน 2 และ3ตามลำดับไป เรื่อยๆ ซึ่งกรณีที่ให้ค่าอคติเข้าใกล้ค่าศูนย์มากที่สุด จะเป็นกรณีที่ดีที่สุด โดยสามารถดูเกณฑ์การให้คะแนนได้ในตารางที่ 4.1 ที่แสดงการให้คะแนนของเกณฑ์อคติและเกณฑ์เกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย

เกณฑ์เกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ใช้เกณฑ์ที่ใช้ประเมินค่าความคลาดเคลื่อนของคุณลักษณะทางสถิติของฝนรายวันที่คำนวณได้ว่ามีการคลาดเคลื่อนมากน้อยเพียงในนั้น ซึ่งหาได้จากสมาการ 4.2

$$RMSE = \beta^2 + \sigma^2 \qquad (4.2)$$

หากค่าเกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่ามากแสดงว่าคุณลักษณะทางสถิติของฝนรายวันที่แบบจำลอง SDSM คำนวณได้ไม่สามารถเลียนแบบคุณลักษณะทางสถิติของฝนรายวันอดีตได้ดีเท่าที่ควร แต่หากค่าเกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าน้อยก็แสดงให้เห็นว่าคุณลักษณะทางสถิติของฝนรายวันที่แบบจำลอง SDSM คำนวณได้สามารถเลียนแบบคุณลักษณะทางสถิติของฝนรายวันอดีตได้ดี

เมื่อคำนวณ ค่า RMSE ดังสมการ 4.2 ครบทุกกรณีแล้ว ก็ทำการให้คะแนนค่าเกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยเช่นเดียวกับเกณฑ์ค่าอคติดังตารางที่ 4.2

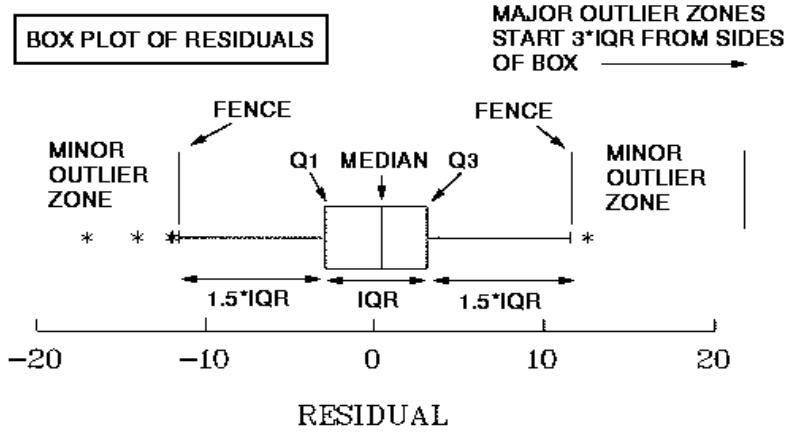
ตารางที่ 4.1

การให้คะแนนเกณฑ์อคติและเกณฑ์เกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย

ความแม่นยำ	น้อยที่สุด	น้อย	ปานกลาง	มาก	มากที่สุด
คะแนน	1	2	3	4	5

คุณลักษณะทางสถิติที่สนใจที่ใช้ในการอธิบายความสามารถของแบบจำลองSDSM ได้แก่ ค่าค่าเฉลี่ย ความแปรปรวน สัมประสิทธิ์ความเบ้ เปอร์เซ็นต์ที่ฝนตก ช่วงเวลาที่ฝนหยุดและตก จำนวนครั้งสัมพัทธ์ที่ฝนหยุดและตก และปริมาณฝนสูงสุด 1-วัน 2-วัน และ 3-วัน โดยแสดงเป็นค่าเฉลี่ยในแต่ละเดือนทั้ง 12 เดือนเพื่อดูคุณลักษณะทางสถิติดังกล่าว ในแต่ละฤดูกาลของลุ่มน้ำ ส่วนค่าปริมาณฝนสูงสุด 1-วัน 2-วัน และ 3-วัน จะแสดงเป็นช่วงเป็นช่วงความถี่ลำดับต่าง เพื่อดูปริมาณของฝนว่ามีปริมาณมากน้อยเพียงใด โดย คุณลักษณะทางสถิติทั้งหมดจะแสดงด้วยการพล็อตรูปกล่อง (box plot) ซึ่ง box pot จะมีลักษณะดังนี้คือ

- box pot แสดงคุณลักษณะต่าง ๆ ของข้อมูลเพิ่มเติม เช่น ค่ากลาง การกระจาย (Spread) การหลุดออกจากการสมมาตร (departure from symmetry) และค่าของตัวอย่างที่หลุดออกจากกลุ่มตัวอย่างไปมาก ๆ (outliers)
- box pot ในส่วนที่เป็นกล่องสี่เหลี่ยมจะแสดงถึง 3 ควอร์ไทล์ ด้านซ้ายหรือด้านล่างของกล่องจะแสดงถึง ควอร์ไทล์แรก (q1) ส่วนด้านบนหรือด้านขวาของกล่องจะแสดงถึง ควอร์ไทล์ที่สาม (q3) เส้นตรงที่อยู่ในกล่องแสดงถึง ควอร์ไทล์ที่ 2 (q2) หรือ มัธยฐาน (median)
- ขอบเขตเส้นตรงที่ลากออกจากล่องทั้งสองด้าน ค่าของข้อมูลที่น้อยที่สุดและที่มากที่สุดที่ตกอยู่ในช่วงข้อมูล 1.5 interquartile (interquartile = q1 ถึง q3)
- จุดที่ปรากฏใน box plot คือค่าของข้อมูลที่หลุดออกนอกกลุ่มมาก ๆ ถ้าจุดที่อยู่ในช่วง 3 interquartile จะเรียกข้อมูลตรงจุดนี้ว่า outlier แต่ถ้าจุดหลุดออกนอกช่วง 3 interquartile จะเรียกจุดนี้ว่า extreme qutlierfy ดังแสดงในภาพที่ 4.3 แสดงคุณลักษณะต่าง ๆ ที่อยู่ใน box plot



ภาพที่ 4.3 คุณลักษณะและตำแหน่งของค่าต่างๆของ box pot