

การเปรียบเทียบวิธีการหาจุดเปลี่ยนแปลงแบบออฟไลน์ในข้อมูลอนุกรมเวลาที่มีความผันแปรไม่ปกติ
ที่มีการแจกแจงปกติ 2 ตัวแปร



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2561
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

COMPARISON OF OFFLINE CHANGE POINT DETECTION METHODS IN BIVARIATE
NORMAL, IRREGULAR TIME SERIES DATA



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Statistics
Department of Statistics
Faculty of Commerce and Accountancy
Chulalongkorn University
Academic Year 2018
Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์

การเปรียบเทียบวิธีการหาจุดเปลี่ยนแปลงแบบออฟไลน์ใน
ข้อมูลอนุกรมเวลาที่มีความผันแปรไม่ปกติที่มีการแจกแจง
ปกติ 2 ตัวแปร

โดย

นายภาวิน เจริญชัยปิยกุล

สาขาวิชา

สถิติ

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

อาจารย์ ดร.อัครินทร์ ไพบูลย์พานิช

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยานิพนธ์ฉบับนี้
เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณะบดีคณะพาณิชยศาสตร์และการ
บัญชี

(ผู้ช่วยศาสตราจารย์ ดร.วิเลิศ ภูริวัชร)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(อาจารย์ ดร.นันท กุลวานิช)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(อาจารย์ ดร.อัครินทร์ ไพบูลย์พานิช)

..... กรรมการ
(อาจารย์ ดร.ณัตติฤดี เจริญรักษ์)

..... กรรมการภายนอกมหาวิทยาลัย
(อาจารย์ ดร.อรุณี กำลั้ง)

ภาววิน เจริญชัยปิยกุล : การเปรียบเทียบวิธีการหาจุดเปลี่ยนแปลงแบบออฟไลน์ใน
ข้อมูลอนุกรมเวลาที่มีความผันแปรไม่ปกติที่มีการแจกแจงปกติ 2 ตัวแปร. (

COMPARISON OF OFFLINE CHANGE POINT DETECTION METHODS IN
BIVARIATE NORMAL, IRREGULAR TIME SERIES DATA) อ.ที่ปรึกษาหลัก : อ. ดร.อัศ

รินทร์ ไพบูลย์พานิช

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการตรวจหาจุด
เปลี่ยนแปลงของวิธี E-Divisive, e-cp30 และ ks-cp30 สำหรับข้อมูลอนุกรมเวลาที่มีความผันแปร
ไม่ปกติที่มีการแจกแจงปกติ 2 ตัวแปรซึ่งมีการเปลี่ยนแปลงในค่าเฉลี่ย ค่าความแปรปรวน หรือค่า
สหสัมพันธ์ โดยทำการจำลองและเปรียบเทียบประสิทธิภาพโดยใช้ค่า Adjusted Rand Index
และเปรียบเทียบจำนวนและตำแหน่งของจุดเปลี่ยนแปลงทั้งสามวิธีในข้อมูลจริง โดยข้อมูลเป็น
ข้อมูลสัญญาณชีพและข้อมูลปริมาณฝุ่นละออง PM2.5 และเปรียบเทียบการเปลี่ยนแปลงของแต่ละ
ช่วงโดยใช้การทดสอบความแตกต่างของค่าเฉลี่ยและความแปรปรวนของสองประชากร

จากการศึกษาพบว่า เมื่อข้อมูลมีขนาดเล็ก ($n = 90$) วิธี e-cp30 และ ks-cp30 มี
ประสิทธิภาพในการตรวจหาจุดเปลี่ยนแปลงมากที่สุด (ค่า Adjusted Rand Index มีค่าเป็น 1)
และพบว่าวิธี E-Divisive มีประสิทธิภาพสูงเฉพาะกรณีที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย (ค่า
Adjusted Rand Index มีค่าเข้าใกล้ 1) เมื่อข้อมูลมีขนาดมากขึ้น ($n = 150$ และ 300) วิธี E-
Divisive มีประสิทธิภาพสูงที่สุดเมื่อข้อมูลมีการเปลี่ยนแปลงค่าเฉลี่ย และวิธี e-cp30 มี
ประสิทธิภาพสูงที่สุดในกรณีอื่น ๆ

การศึกษากับข้อมูลจริงซึ่งเป็นข้อมูลสัญญาณชีพ และข้อมูลปริมาณฝุ่นละออง
PM2.5 ซึ่งทำการตรวจหาจุดเปลี่ยนแปลงทั้งหมดในคราวเดียวและแบ่งข้อมูลออกเป็น 3 ส่วนก่อน
แล้วจึงนำแต่ละส่วนมาตรวจหาจุดเปลี่ยนแปลง พบว่า วิธี E-Divisive และ ks-cp30 พบจำนวนจุด
และตำแหน่งของจุดเปลี่ยนแปลงที่ใกล้เคียงกัน และการแบ่งข้อมูลออกเป็นช่วงย่อยก่อนจะมี
ประสิทธิภาพมากกว่าการตรวจข้อมูลทั้งหมดในคราวเดียวกัน

สาขาวิชา สถิติ

ลายมือชื่อนิสิต

ปีการศึกษา 2561

ลายมือชื่อ อ.ที่ปรึกษาหลัก

6081566926 : MAJOR STATISTICS

KEYWORD: Change point detection, change point, bivariate normal time series,
offline data

Papawin Charoenchaiyikul : COMPARISON OF OFFLINE CHANGE POINT
DETECTION METHODS IN BIVARIATE NORMAL, IRREGULAR TIME SERIES DATA.

Advisor: AKARIN PHAIBULPANICH, Ph.D.

The objective of this research is to compare the performances of 3 change point detection methods: “E-Divisive”, “e-cp3o” and “ks-cp3o”, for bivariate normal irregular time series data which has a change in mean, variance or correlation. Their performances are compared using Adjusted Rand Index. The methods are also applied to two real data sets (vital sign data and concentration of PM2.5 data) and compared.

The results show that: 1) when data is small ($n = 90$), e-cp3o and ks-cp3o performed the best (Adjusted Rand Indexes are 1), E-Divisive's Adjusted Rand Index is almost 1 when data has a change in mean, 2) When data is large ($n = 150$ and 300), e-cp3o and ks-cp3o are less efficient, while E-Divisive performed the best when data has a change in mean.

The study of real data by detecting change points in the whole data at once and dividing the data into 3 parts then detect the change points in each part. The results show that E-Divisive and ks-cp3o detect similarly in both the number and the position of change points and the performances are better when dividing the data into smaller intervals before detecting the change points.

Field of Study: Statistics

Student's Signature

Academic Year: 2018

Advisor's Signature

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลงได้ด้วยความรู้ และความอนุเคราะห์ช่วยเหลืออย่างดียิ่งจาก อาจารย์ ดร.อัครินทร์ ไพบูลย์พานิช อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้โอกาสผู้วิจัยได้เป็นลูกศิษย์ใน ที่ปรึกษา คอยให้ความรู้คำแนะนำ ตลอดจนชี้แนะแนวทางในการศึกษาแก่ผู้วิจัย อีกทั้งยังช่วยแก้ไข ข้อบกพร่องต่างๆด้วยความเอาใจใส่ จนกระทั่งวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี ผู้วิจัยขอกราบ ขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ อาจารย์ ดร.นัท กุลวานิช ประธานกรรมการสอบวิทยานิพนธ์ อาจารย์ ดร.ณตติฤดี เจริญรักษ์ กรรมการสอบวิทยานิพนธ์ และอาจารย์ ดร. อรุณี กำลัง กรรมการ ผู้ทรงคุณวุฒิภายนอก ที่ได้กรุณาสละเวลามาดูงานแก้ไขข้อบกพร่อง ในวิทยานิพนธ์ฉบับนี้ ตลอดจน ให้คำแนะนำที่เป็นประโยชน์แก่ผู้วิจัยที่จะช่วยให้วิทยานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น อีกทั้งคุณอาจารย์ประ จาภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัยทุกท่านที่ได้ให้โอกาส ทางการศึกษา และประสิทธิประสาทวิชาความรู้ทางด้านสถิติให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษาใน ระดับปริญญาโท

สุดท้ายนี้ผู้วิจัยขอกราบขอบพระคุณบิดา มารดา และครอบครัว ที่คอยสนับสนุน เป็นกำลังใจ ให้กับผู้วิจัยในการศึกษาในระดับปริญญาโท ตลอดจนการทำวิทยานิพนธ์ฉบับนี้จนสำเร็จลุล่วงไปได้ ด้วยดี

คุณประโยชน์ และความดีใดๆอันพึงมีจากวิทยานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบและอุทิศ ให้แก่ บิดา มารดา ครอบครัว และคุณอาจารย์ประจาภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัยทุกท่าน ที่ได้ให้ความรู้ และวางรากฐานทางการศึกษาให้แก่ผู้วิจัยต่อไปใน อนาคต

ปภาวิน เจริญชัยปิยกุล

สารบัญ

	หน้า
.....	ค
บทคัดย่อภาษาไทย.....	ค
.....	ง
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ญ
สารบัญรูปภาพ.....	ฎ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์การวิจัย.....	3
1.3 ประโยชน์ที่คาดว่าจะได้รับ.....	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	4
2.1 Change Point Procedure via Pruned Objectives (cp3o).....	4
2.1.1 Energy statistic.....	7
2.1.2 \mathcal{A} -Distance.....	7
2.2 E-divisive.....	8
2.3 Rand index.....	9
2.4 การทดสอบความแตกต่างระหว่างค่าเฉลี่ยสองประชากร	10
2.5 การทดสอบความแตกต่างระหว่างค่าความแปรปรวนสองประชากร.....	12
บทที่ 3 ขอบเขตการวิจัย	14

3.1	เงื่อนไขที่ทำการศึกษา.....	14
3.2	วิธีดำเนินการวิจัย.....	14
3.3	ข้อจำกัดในงานวิจัย.....	23
บทที่ 4 ผลการวิจัย.....		24
4.1	ผลการเปรียบเทียบวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o ในส่วนข้อมูลจำลอง.....	25
4.1.1	ผลการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธีE-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 90.....	25
4.1.2	ผลการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธีE-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 150.....	26
4.1.3	ผลการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธีE-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 300.....	27
4.1.4	ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 90	28
4.1.5	ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 150	29
4.1.6	ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 300	30
4.1.7	ผลการเปรียบเทียบประสิทธิภาพและจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive และ e-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 90	31
4.1.8	ผลการเปรียบเทียบประสิทธิภาพและจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive และ e-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 150.....	32
4.1.9	ผลการเปรียบเทียบประสิทธิภาพและจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive และ e-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 300.....	33
4.2	ผลการเปรียบเทียบวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o ในส่วนข้อมูลจริง.....	34

4.2.1 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงของข้อมูลสัญญาณชีพที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o โดยทำการตรวจหาข้อมูลทั้งหมดในคราวเดียว	34
4.2.2 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงของข้อมูลสัญญาณชีพที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o โดยทำการแบ่งข้อมูลออกเป็น 3 ส่วนก่อน... 43	43
4.2.3 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงของข้อมูลปริมาณฝุ่นละออง PM2.5 ที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o โดยทำการตรวจหาข้อมูลทั้งหมดในคราวเดียว	46
4.2.4 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงของข้อมูลปริมาณฝุ่นละออง PM2.5 ที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o โดยทำการแบ่งข้อมูลออกเป็น 3 ส่วนก่อน.....	54
บทที่ 5 สรุปผลวิจัยและข้อเสนอแนะ	57
5.1 สรุปผลการวิจัย	57
5.1.1 ผลการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 90.....	57
5.1.2 ผลการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 150.....	58
5.1.3 ผลการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 300.....	59
5.1.4 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 90	60
5.1.5 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 150	61
5.1.6 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 300	62
5.1.7 ผลการเปรียบเทียบวิธีการตรวจหาจุดเปลี่ยนแปลงวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อนำไปตรวจหาจุดเปลี่ยนแปลงในข้อมูลจริง	63

5.2 ข้อเสนอแนะ	65
บรรณานุกรม.....	66
ภาคผนวก.....	68
ประวัติผู้เขียน.....	141



สารบัญตาราง

หน้า

ตารางที่ 1 ค่าเฉลี่ย Adjusted Rand Index ของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่างเท่ากับ 90	25
ตารางที่ 2 ค่าเฉลี่ย Adjusted Rand Index ของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่างเท่ากับ 150....	26
ตารางที่ 3 ค่าเฉลี่ย Adjusted Rand Index ของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่างเท่ากับ 300....	27
ตารางที่ 4 ค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงที่พบของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่าง 90	28
ตารางที่ 5 ค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงที่พบของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่าง 150	29
ตารางที่ 6 ค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงที่พบของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่าง 300	30
ตารางที่ 7 ค่าเฉลี่ย Adjusted Rand Index และค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงของวิธี E-Divisive และ e-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 90	31
ตารางที่ 8 ค่าเฉลี่ย Adjusted Rand Index และค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงของวิธี E-Divisive และ e-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 150	32
ตารางที่ 9 ค่าเฉลี่ย Adjusted Rand Index และค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงของวิธี E-Divisive และ e-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 300	33
ตารางที่ 10 ค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลงเมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive	35
ตารางที่ 11 ค่า p-value การทดสอบสมมติฐานการเปลี่ยนแปลงค่าเฉลี่ย และค่าความแปรปรวนในแต่ละช่วง ที่ระดับนัยสำคัญ 0.05 เมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive	37
ตารางที่ 12 ค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลงเมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี e-cp3o	38
ตารางที่ 13 ค่า p-value การทดสอบสมมติฐานการเปลี่ยนแปลงค่าเฉลี่ย และค่าความแปรปรวนในแต่ละช่วง ที่ระดับนัยสำคัญ 0.05 เมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี e-cp3o	39
ตารางที่ 14 ค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลงเมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี ks-cp3o	40

ตารางที่ 15 ค่า p-value การทดสอบสมมติฐานการเปลี่ยนแปลงค่าเฉลี่ย และค่าความแปรปรวนในแต่ละช่วง ที่ระดับนัยสำคัญ 0.05 เมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี ks-cp3o	42
ตารางที่ 16 ค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลงเมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive	47
ตารางที่ 17 ค่า p-value การทดสอบสมมติฐานการเปลี่ยนแปลงค่าเฉลี่ย และค่าความแปรปรวนในแต่ละช่วง ที่ระดับนัยสำคัญ 0.05 เมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive	48
ตารางที่ 18 ค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลงเมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี e-cp3o	50
ตารางที่ 19 ค่า p-value การทดสอบสมมติฐานการเปลี่ยนแปลงค่าเฉลี่ย และค่าความแปรปรวนในแต่ละช่วง ที่ระดับนัยสำคัญ 0.05 เมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี e-cp3o	51
ตารางที่ 20 ค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลงเมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี ks-cp3o	52
ตารางที่ 21 ค่า p-value การทดสอบสมมติฐานการเปลี่ยนแปลงค่าเฉลี่ย และค่าความแปรปรวนในแต่ละช่วง ที่ระดับนัยสำคัญ 0.05 เมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี ks-cp3o	53
ตารางที่ 22 ตารางแสดงผลการเปรียบเทียบประสิทธิภาพวิธีการตรวจหาจุดเปลี่ยนแปลงเมื่อข้อมูลมีขนาด 90	58
ตารางที่ 23 ตารางแสดงผลการเปรียบเทียบประสิทธิภาพวิธีการตรวจหาจุดเปลี่ยนแปลงเมื่อข้อมูลมีขนาด 150	59
ตารางที่ 24 ตารางแสดงผลการเปรียบเทียบประสิทธิภาพวิธีการตรวจหาจุดเปลี่ยนแปลงเมื่อข้อมูลมีขนาด 300	60
ตารางที่ 25 ตารางแสดงผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อข้อมูลมีขนาด 90	61
ตารางที่ 26 ตารางแสดงผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อข้อมูลมีขนาด 150	62
ตารางที่ 27 ตารางแสดงผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อข้อมูลมีขนาด 300	63

สารบัญรูปภาพ

	หน้า
รูปที่ 1 แผนผังการวิเคราะห์การเดิน (gait analysis)	2
รูปที่ 2 ลำดับข้อมูล Z_1, Z_2, \dots, Z_t มีจุดเปลี่ยนแปลง k จุดที่ตำแหน่ง $t_1 < t_2 < \dots < t_{k-1} < t_k$	4
รูปที่ 3 ประมาณตำแหน่ง t_1, t_2, \dots, t_k ที่ตำแหน่ง $\tau_1 < \tau_2 < \dots < \tau_{k-1} < \tau_k$ และคำนวณค่าความเหมาะสม	5
รูปที่ 4 การคำนวณเพื่อหาตำแหน่งของจุดแบ่งกันที่ K	6
รูปที่ 5 ตัวอย่างข้อมูลที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย	15
รูปที่ 6 ตัวอย่างข้อมูลที่ข้อมูลมีการเปลี่ยนแปลงในค่าความแปรปรวน	16
รูปที่ 7 ตัวอย่างข้อมูลที่ข้อมูลมีการเปลี่ยนแปลงในค่าสหสัมพันธ์	17
รูปที่ 8 ตัวอย่างข้อมูลที่ข้อมูลมีการเปลี่ยนแปลงในค่าสหสัมพันธ์	18
รูปที่ 9 ตัวอย่างข้อมูลที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ยและค่าความแปรปรวน	19
รูปที่ 10 ตัวอย่างข้อมูลที่ข้อมูลมีการเปลี่ยนแปลงในค่าความแปรปรวนและค่าสหสัมพันธ์	20
รูปที่ 11 ตัวอย่างข้อมูลที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ยและค่าสหสัมพันธ์	21
รูปที่ 12 ตัวอย่างข้อมูลที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์	22
รูปที่ 13 แผนภูมิแสดงค่าเฉลี่ย Adjusted Rand Index ของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่างเท่ากับ 90	25
รูปที่ 14 แผนภูมิแสดงค่าเฉลี่ย Adjusted Rand Index ของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่างเท่ากับ 150	26
รูปที่ 15 แผนภูมิแสดงค่าเฉลี่ย Adjusted Rand Index ของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่างเท่ากับ 300	27
รูปที่ 16 แผนภูมิแสดงค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงที่พบของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่าง 90	28
รูปที่ 17 แผนภูมิแสดงค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงที่พบของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่าง 150	29

รูปที่ 18 แผนภูมิแสดงค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงที่พบของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่าง 300 30

รูปที่ 19 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive โดยตรวจข้อมูลทั้งหมด..... 34

รูปที่ 20 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี e-cp3o โดยตรวจข้อมูลทั้งหมด..... 37

รูปที่ 21 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี ks-cp3o โดยตรวจข้อมูลทั้งหมด 39

รูปที่ 22 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive โดยทำการแบ่งข้อมูลก่อน .. 43

รูปที่ 23 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี e-cp3o โดยทำการแบ่งข้อมูลก่อน 44

รูปที่ 24 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี ks-cp3o โดยทำการแบ่งข้อมูลก่อน 44

รูปที่ 25 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive โดยตรวจข้อมูลทั้งหมด..... 46

รูปที่ 26 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี e-cp3o โดยตรวจข้อมูลทั้งหมด..... 49

รูปที่ 27 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี ks-cp3o โดยตรวจข้อมูลทั้งหมด 51

รูปที่ 28 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive โดยทำการแบ่งข้อมูลก่อน .. 54

รูปที่ 29 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี e-cp3o โดยทำการแบ่งข้อมูลก่อน 54

รูปที่ 30 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี ks-cp3o โดยทำการแบ่งข้อมูลก่อน 55

บทที่ 1

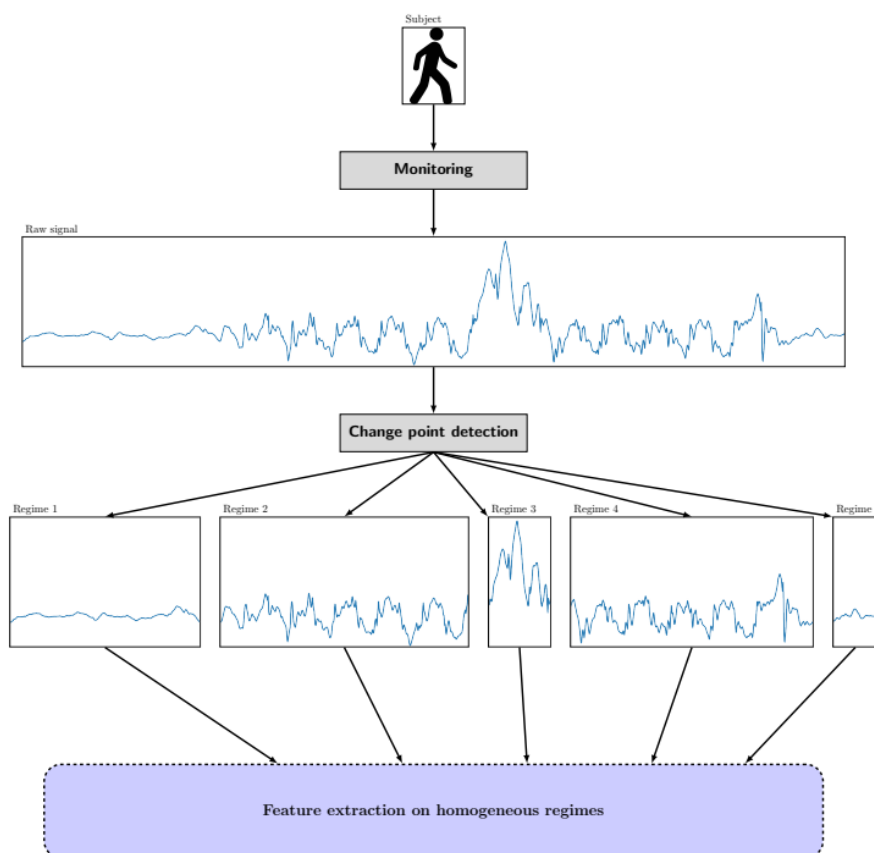
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การตรวจหาจุดเปลี่ยนแปลง (Change point detection) คือ กระบวนการในการหาจุดเปลี่ยนแปลงภายในรูปแบบของสัญญาณหรืออนุกรมเวลา (Truong, Oudre, & Vayatis, 2018) งานวิจัยเริ่มแรกที่เกี่ยวข้องกับการตรวจหาจุดเปลี่ยนแปลงมีเป้าหมายในการตรวจหาการเปลี่ยนแปลงของค่าเฉลี่ยสำหรับการควบคุมคุณภาพอุตสาหกรรม ซึ่งปัญหาการตรวจหาจุดเปลี่ยนแปลงไม่เพียงเข้ามามีบทบาทที่สำคัญในทางประมวลผลสัญญาณและทางสถิติ แต่ยังมีบทบาทในด้านต่าง ๆ มากมาย เช่น การวิเคราะห์ทางการเงิน ชีวสารสนเทศ (bioinformatics) อดุณิยวิทยา (Wang, Killick, & Fu, 2014) เป็นต้น

อีกหนึ่งตัวอย่างการนำการตรวจหาจุดเปลี่ยนแปลงไปใช้ คือ การวิเคราะห์การเดิน (Truong et al., 2018) โดยทำการเก็บข้อมูลการเคลื่อนไหวของบุคคลในขณะที่ทำกิจกรรมง่าย ๆ เช่น การเดิน การวิ่ง หรือการยืน โดยใช้เครื่องมือวัดการเร่ง (accelerometer) และเครื่องมือวัดการหมุนวน (gyroscope) ได้แบ่งสัญญาณที่แบ่งออกเป็นส่วน ๆ ดังแสดงในรูปที่ 1 โดยแต่ละส่วนนั้นจะบ่งบอกถึงลักษณะที่แตกต่างกัน

การตรวจหาจุดเปลี่ยนแปลงนั้นสามารถแบ่งออกได้เป็น 2 แบบ คือ วิธีการแบบออนไลน์ (online method) ที่มีเป้าหมายในการตรวจหาการเปลี่ยนแปลงทันที และ-วิธีการแบบออฟไลน์ (offline method) ที่จะทำการตรวจสอบการเปลี่ยนแปลงหลังจากที่ได้รับข้อมูลทั้งหมดแล้ว ซึ่งวิธีการตรวจหาจุดเปลี่ยนแปลงนั้นมีหลายวิธี (Truong et al., 2018) โดยแบ่งได้ตามประเภทของฟังก์ชันต้นทุน (cost function) คือค่าที่นำไปวัดความเหมาะสมในการประมาณตำแหน่งจุดเปลี่ยนแปลง วิธีการค้นหา (search method) คือวิธีการประมาณจำนวนจุดเปลี่ยนแปลง และการเพิ่มเงื่อนไขบนจำนวนจุดที่เปลี่ยน (constraint on the number of change points)



รูปที่ 1 แผนผังการวิเคราะห์การเดิน (gait analysis)

ในการแบ่งวิธีการตรวจหาจุดเปลี่ยนแปลงโดยดูจากฟังก์ชันต้นทุน สามารถแบ่งออกได้เป็นแบบ parametric ที่สนใจในการเปลี่ยนแปลงของเวกเตอร์พารามิเตอร์ และแบบ nonparametric ที่อาจจะดีกว่าวิธีการแบบ parametric ในกรณีที่ข้อมูลไม่สอดคล้องกับเงื่อนไขบางประการ หากดูจากวิธีการค้นหา จะสามารถแบ่งออกได้เป็น แบบที่ดีที่สุด (optimal) ที่จะคำนวณผลลัพธ์ของจุดที่เปลี่ยนแปลงทุกแบบที่เป็นไปได้ และให้ค่าที่ดีที่สุด และแบบการประมาณ (approximate) ที่ประมาณผลลัพธ์ออกมาในกรณีที่วิธีที่ดีที่สุดไม่สามารถคำนวณออกมาได้ และการเพิ่มเงื่อนไขจำนวนจุดที่เปลี่ยนที่จะขึ้นกับว่าทราบจำนวนจุดที่มีการเปลี่ยนแปลงมาก่อนหรือไม่

โดยในงานวิจัยนี้ ผู้วิจัยมีความสนใจในวิธี E-Divisive (Matteson & James, 2014) ซึ่งเป็นวิธีการแบบ nonparametric โดยจะทำการตรวจหาทั้งจำนวนจุดที่มีการเปลี่ยนแปลง และจุดที่การเปลี่ยนแปลงเกิดขึ้น โดยการคำนวณค่าวัดความแตกต่าง (divergence measure) จาก Energy statistic (Szekely & Rizzo, 2005) ในข้อมูลที่ถูกแบ่งออกเป็นผลแบ่งกัน และทดสอบสมมติฐานเพื่อตรวจสอบว่ามีจุดเปลี่ยนแปลงอีกหรือไม่ และ วิธี cp3o (Change Point Procedure via Pruned

Objectives) (Zhang, James, & Matteson, 2017) เป็นขั้นตอนการค้นหาจุดเปลี่ยนแปลงและนำจุดเปลี่ยนแปลงที่ได้ไปทดสอบความเหมาะสมโดยใช้ Energy statistic และ Kolmogorov-Smirnov statistic ซึ่งเรียกว่า e-cp3o และ ks-cp3o ตามลำดับ

จากที่กล่าวมาข้างต้น ผู้วิจัยจึงมีความสนใจที่จะศึกษาและทำการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงแบบ nonparametric 3 วิธีข้างต้น เนื่องจากวิธีทั้ง 3 วิธีเป็นวิธีที่มีเงื่อนไขในการใช้งานและข้อสมมติฐานน้อย และยังไม่มียานวิจัยที่เปรียบเทียบประสิทธิภาพในกรณีของข้อมูลที่เป็นออฟไลน์ที่มีการแจกแจงปกติ 2 ตัวแปร โดยผู้วิจัยจะเปรียบเทียบประสิทธิภาพการตรวจหาการเปลี่ยนแปลงในส่วนของข้อมูลจำลองที่มีการเปลี่ยนแปลงในค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ และในส่วนของข้อมูลจริงซึ่งเป็นข้อมูลของมหาวิทยาลัยควีนส์แลนด์ (The university of Queensland) ในกรณีตัวอย่างที่ 2 (Liu, Görges, & Jenkins, 2012) โดยมีตัวแปรที่สนใจคือ อัตราการเต้นของหัวใจ (Heart rate) และอัตราการเต้นของชีพจร (Pulse) ซึ่งมีขนาด 900 วินาที (นาทีที่ 0.12 ถึง นาทีที่ 15.11) และข้อมูลปริมาณฝุ่นละออง PM2.5 จาก BERKELEY EARTH ซึ่งมีขนาด 720 จุด (ข้อมูล ณ วันที่ 1 พฤศจิกายน 2561 ถึง 30 พฤศจิกายน 2561) ซึ่งมีตัวแปรที่สนใจคือ ความหนาแน่นของปริมาณฝุ่นละออง PM2.5 และอุณหภูมิและเพื่อเป็นทางเลือกให้ผู้สนใจสามารถนำเอาแต่ละวิธีไปประยุกต์ใช้กับการตรวจหาจุดเปลี่ยนแปลงในชีวิตจริงได้

1.2 วัตถุประสงค์การวิจัย

1. เพื่อเปรียบเทียบประสิทธิภาพในการตรวจหาจุดเปลี่ยนแปลงของวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลอนุกรมเวลาที่มีความผันแปรไม่ปกติที่มีการแจกแจงปกติ 2 ตัวแปรมีการเปลี่ยนแปลงในค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์

2. เพื่อเปรียบเทียบจำนวนและตำแหน่งจุดเปลี่ยนแปลงที่พบจากการตรวจหาจุดเปลี่ยนแปลงของวิธี E-Divisive, e-cp3o และ ks-cp3o ในส่วนของข้อมูลจริงเมื่อข้อมูลเป็นข้อมูลสัญญาณชีพ ซึ่งมีตัวแปรที่สนใจคือ อัตราการเต้นของหัวใจ (Heart rate) และอัตราการเต้นของชีพจร (Pulse) และข้อมูลปริมาณฝุ่นละออง PM2.5 ซึ่งมีตัวแปรที่สนใจคือ ความหนาแน่นของปริมาณฝุ่นละออง PM2.5 และอุณหภูมิ

1.3 ประโยชน์ที่คาดว่าจะได้รับ

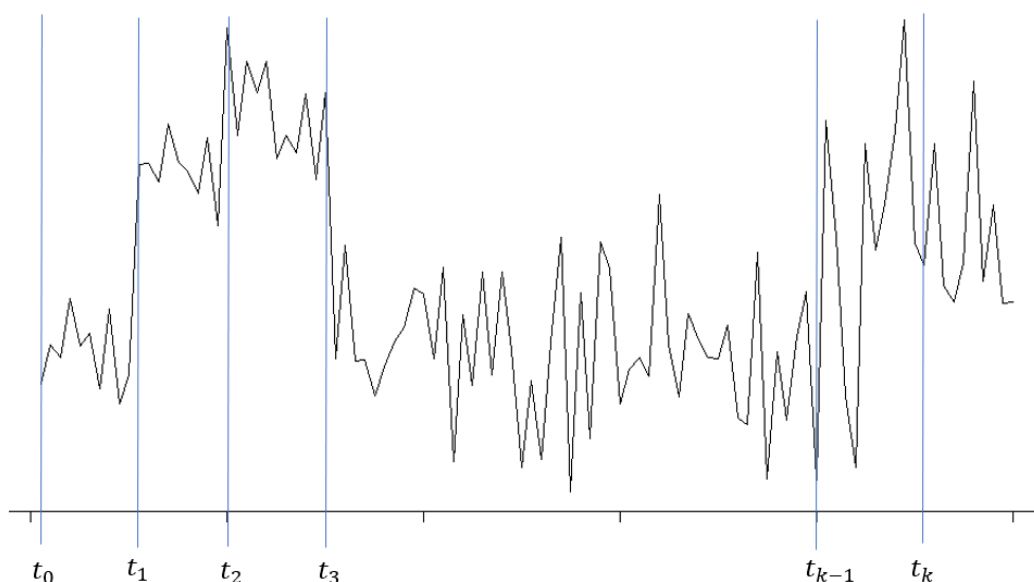
สามารถนำไปเป็นแนวทางในการเลือกใช้วิธีการตรวจหาจุดเปลี่ยนแปลงเพื่อตรวจหาการเปลี่ยนแปลงในค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ ผู้วิจัยจะกล่าวถึงแนวคิดทฤษฎีต่าง ๆ ที่เกี่ยวข้องกับการเปรียบเทียบประสิทธิภาพวิธีการตรวจหาจุดเปลี่ยนแปลง โดยผู้วิจัยได้แบ่งการนำเสนอออกเป็น 5 ส่วนหลัก คือ วิธี cp3o, วิธี E-Divisive, ค่า Rand Index, การทดสอบความแตกต่างค่าเฉลี่ยของสองประชากร และการทดสอบความแตกต่างค่าความแปรปรวนระหว่างสองประชากร โดยมีรายละเอียดดังต่อไปนี้

2.1 Change Point Procedure via Pruned Objectives (cp3o)

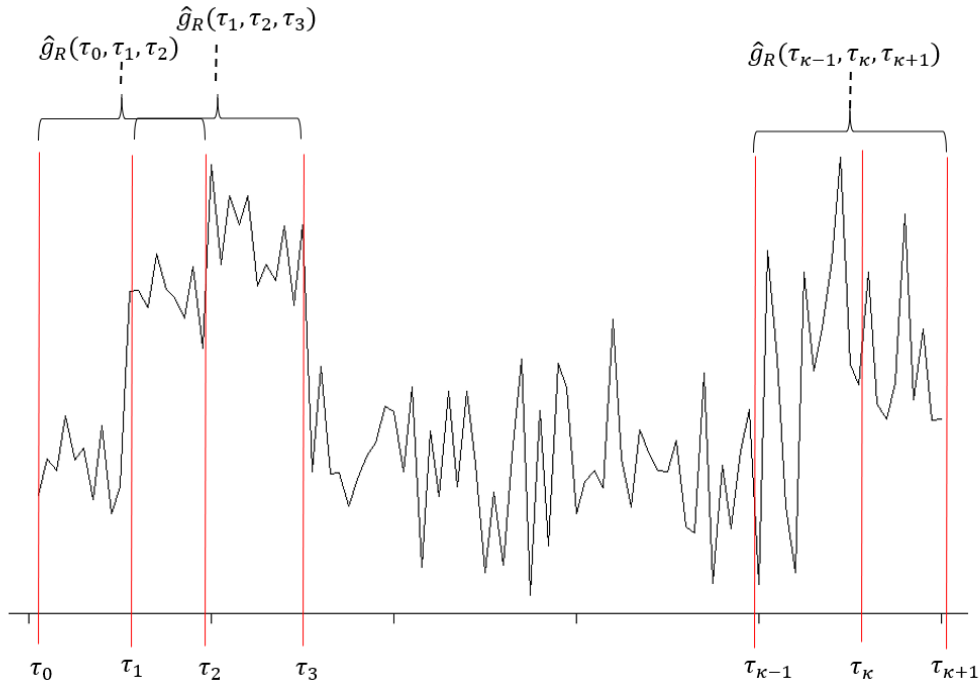


รูปที่ 2 ลำดับข้อมูล Z_1, Z_2, \dots, Z_T มีจุดเปลี่ยนแปลง k จุดที่ตำแหน่ง $t_1 < t_2 < \dots < t_{k-1} < t_k$

ให้ $Z_1, Z_2, \dots, Z_T \in \mathbb{R}^d$ เป็นลำดับที่มีความยาว T ของตัวแปรสุ่มที่ถูกจัดเรียงด้วยเวลาใน d มิติที่อิสระต่อกัน และให้ k เป็นจำนวนจุดที่มีการเปลี่ยนแปลงโดยมีจุดเปลี่ยนแปลงที่เวลา $1 = t_0 < t_1 < \dots < t_k < t_{k+1} = T + 1$ เมื่อ $Z_i \sim F_j$ แจกแจงแบบอิสระเหมือนกันสำหรับ $t_j \leq i < t_{j+1}$ และ $F_j \neq F_{j+1}$ สำหรับ $0 \leq j \leq k$ และ $\omega \geq 1$ เป็นขอบเขตล่างของระยะห่างระหว่างจุดเปลี่ยน จากนั้นแบ่งลำดับ Z_1, Z_2, \dots, Z_T ออกเป็น k ผลแบ่งกัน โดยพิจารณาคุณภาพของการแบ่งจากค่าความเหมาะสมเชิงประจักษ์ (empirical goodness of fit metric)

$$\sum_{j=1}^{\kappa} \hat{g}_R(\tau_{j-1}, \tau_j, \tau_{j+1})$$

เมื่อ $\hat{g}_R(a, b, c) = \hat{R}(Z_a^{b-1}, Z_b^{c-1})$ โดยที่ $Z_a^b = \{Z_i\}_{i=a}^b$ สำหรับ $a < b < c$ และ $\hat{R}(\cdot, \cdot)$ คือ ค่าที่วัดความต่างระหว่างการแจกแจงของ 2 ตัวแปรสุ่ม



รูปที่ 3 ประมาณตำแหน่ง t_1, t_2, \dots, t_k ที่ตำแหน่ง $\tau_1 < \tau_2 < \dots < \tau_{\kappa-1} < \tau_{\kappa}$ และคำนวณค่าความเหมาะสม

ซึ่งค่าที่มากที่สุดของค่าความเหมาะสมเชิงประจักษ์ของการแบ่งลำดับความยาว T ออกเป็น κ ผลแบ่งกัน ที่มีจุดเปลี่ยนแปลง k จุดคือ

$$\hat{G}_T(\kappa, \omega) = \max_{\substack{\tau_1, \tau_2, \dots, \tau_{\kappa} \\ \tau_i + \omega \leq \tau_{i+1}, i < \kappa \\ \tau_i \in \{1 + \omega, \dots, T - \omega + 1\}}} \sum_{j=1}^{\kappa} \hat{g}_R(\tau_{j-1}, \tau_j, \tau_{j+1})$$

ในการประมาณค่า κ และ $\{\tau_i\}_{i=1}^{\kappa}$ สำหรับจำนวนจุดเปลี่ยนแปลง k และตำแหน่งที่มีการเปลี่ยนแปลง $\{t_i\}_{i=1}^k$ ต้องคำนวณ $\hat{G}_T(\kappa, \omega)$ สำหรับ $1 \leq \kappa \leq K$ เมื่อ $K \geq k$ คือขอบเขตบนของ k และ $\hat{G}_T(\kappa, \omega)$ สามารถคำนวณได้โดย

ให้ $H_t(\kappa, \omega, \tau) = \tilde{G}_{\tau-1}(\kappa - 1, \omega) + \hat{g}_R(A_{\tau-1}(\kappa - 1), \tau, t)$ โดยในกระบวนการรอบที่ κ สำหรับแต่ละลำดับย่อย $\{Z_i\}_{i=1}^t$ เมื่อ $1 \leq t \leq T$ ให้

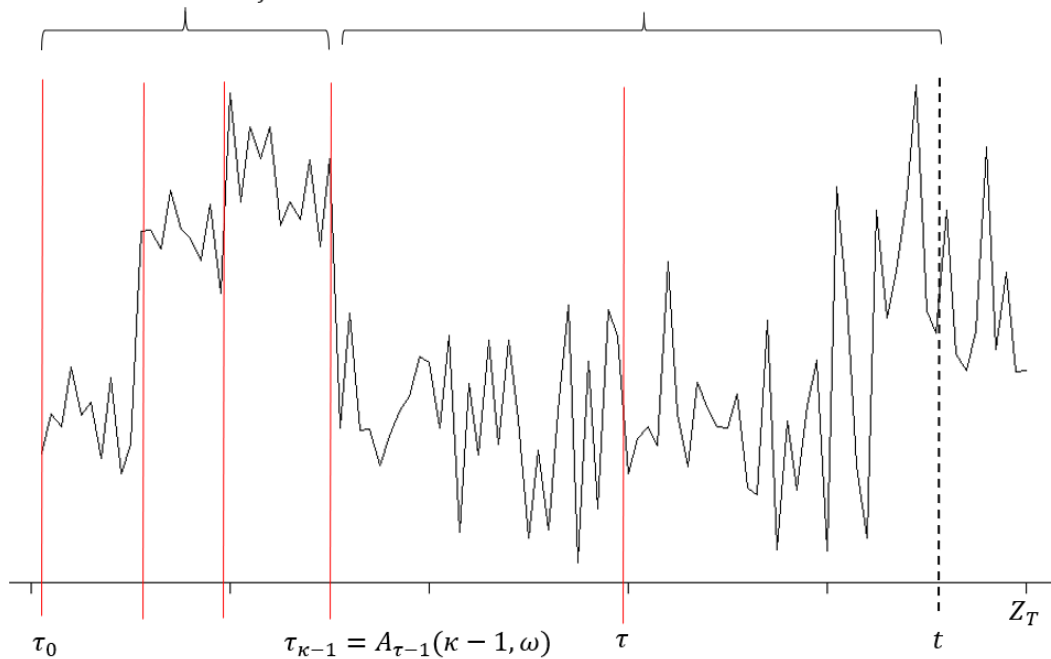
$$A_t(\kappa, \omega) = \operatorname{argmax}_{\tau \in \{1 + \kappa^* \omega, \dots, t - \omega + 1\}} H_t(\kappa, \omega, \tau)$$

$$\tilde{G}_t(\kappa, \omega) = \max_{\tau \in \{1 + \kappa^* \omega, \dots, t - \omega + 1\}} H_t(\kappa, \omega, \tau)$$

โดยที่ $\tilde{G}_t(\kappa, \omega)$ คือ ค่าประมาณที่ดีที่สุดของค่าความเหมาะสมของการแบ่งลำดับความยาว t ที่มีจุดแบ่งกัน κ จุด

$A_t(\kappa, \omega)$ คือ ตำแหน่งของจุดแบ่งกันที่ κ

$$\tilde{G}_{\tau-1}(\kappa - 1, \omega) = \sum_{j=1}^{\kappa-1} \hat{g}_R(\tau_{j-1}, \tau_j, \tau_{j+1}) \quad \hat{g}_R(\tau_{\kappa-1}, \tau, t)$$



รูปที่ 4 การคำนวณเพื่อหาตำแหน่งของจุดแบ่งกันที่ κ

ในกระบวนการรอบที่ κ การคำนวณ $A_t(\kappa, \omega)$ และ $\tilde{G}_t(\kappa, \omega)$ จะต้องหาจุดแบ่ง κ ที่ดีที่สุดของ $\{Z_i\}_{i=1}^t$ จาก $\{1 + \kappa^* \omega, \dots, t - \omega + 1\}$ ดังนั้นเพื่อที่จะลดการคำนวณจึงลดเซตที่ต้องคำนวณโดย สำหรับรอบแรกในการคำนวณ ให้

$$S_t(1, \omega) = \{1 + \omega, \dots, t - \omega + 1\}$$

ซึ่งเป้นเซตในการคำนวณที่ใหญ่ที่สุด และสำหรับรอบที่ $\kappa + 1$ ให้ $S_t(\kappa + 1, \omega)$ คือผลจากการตัดเซตค้นหา $S_t(\kappa, \omega)$ จากกระบวนการรอบก่อนหน้า นั่นคือ

$$S_t(\kappa + 1, \omega) = \{\tau \in S_t(\kappa, \omega) : H_t(\kappa + 1, \omega, \tau) \geq H_t(\kappa + 1, \omega, t - \omega + 1)\}$$

โดยในอสมการข้างต้นเป็นการเปรียบเทียบความเหมาะสมของการแบ่งลำดับย่อยความยาว t ออกเป็น $\kappa + 1$ ผลแบ่งกัน 2 แบบ โดยที่อันหนึ่งมีจุดเปลี่ยนแปลงสุดท้ายที่จุด τ และอีกอันหนึ่งที่จุด $t - \omega + 1$ ซึ่งถ้าค่าที่จุด τ ต่ำกว่า จะแสดงว่าจุด τ ไม่ใช่จุดที่มีการเปลี่ยนแปลง

2.1.1 Energy statistic

สำหรับ $\alpha \in (0, 2)$ ให้ $X_n = \{X_i\}_{i=1}^n$ และ $Y_m = \{Y_j\}_{j=1}^m$ เป็นตัวอย่างที่มีการแจกแจงความน่าจะเป็น F_X และ F_Y ซึ่งอิสระต่อกันตามลำดับ ค่าวัดความต่างระหว่างการแจกแจงคือ

$$\mathcal{E}(X, Y|\alpha) = 2E|X - Y|^\alpha - E|X - X|^\alpha - E|Y - Y|^\alpha$$

ค่าเชิงประจักษ์ของ $\mathcal{E}(X, Y|\alpha)$ คือ

$$\hat{\mathcal{E}}(X_n, Y_m|\alpha) = \frac{2}{mn} \sum_{i=1}^n \sum_{j=1}^m |x_i - y_j|^\alpha - \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} |x_i - x_j|^\alpha - \binom{m}{2}^{-1} \sum_{1 \leq i < j \leq m} |y_i - y_j|^\alpha$$

ให้ γ คือ สัดส่วนของข้อมูลจาก F_X ในลิมิตเมื่อ $\min(n, m) \rightarrow \infty$ จะได้ว่าค่าความแตกต่างคือ

$$R(X, Y|\alpha) = \gamma(1 - \gamma)\mathcal{E}(X, Y|\alpha)$$

$$\hat{R}(X_n, Y_m|\alpha) = \frac{mn}{(m + n)^2} \hat{\mathcal{E}}(X_n, Y_m|\alpha)$$

2.1.2 \mathcal{A} -Distance

ให้ \mathcal{A} เป็นเซตของ measurable set โดย \mathcal{A} -Distance ได้ถูกนิยามไว้ว่า

$$d_{\mathcal{A}}(F_X, F_Y) = 2 \sup_{A \in \mathcal{A}} |F_X(A) - F_Y(A)|$$

และ \mathcal{A} -Distance เชิงประจักษ์คือ

$$\hat{d}(X_n, Y_m|\mathcal{A}) = 2 \sup_{A \in \mathcal{A}} \left| \frac{|x_n \cap A|}{n} - \frac{|y_m \cap A|}{m} \right|$$

ให้ γ คือ สัดส่วนของข้อมูลจาก F_X ในลิมิตเมื่อ $\min(n, m) \rightarrow \infty$ จะได้ว่าค่าความแตกต่างคือ

$$R(X, Y|\alpha) = \gamma(1 - \gamma)d_{\mathcal{A}}(F_X, F_Y)$$

$$\hat{R}(X_n, Y_m|\mathcal{A}) = \frac{mn}{(m+n)^2} \hat{d}(X_n, Y_m|\mathcal{A})$$

โดยสำหรับ $\mathcal{A} = \{(-\infty, r) | r \in \mathbb{R}\}$ จะได้ว่า $\hat{d}(X_n, Y_m|\mathcal{A})$ คือ Kolmogorov-Smirnov statistic

โดยถ้าหากค่าวัดความต่างที่ใช้คือ Energy statistics และ Kolmogorov-Smirnov statistics จะเรียกวินี้ว่า “e-cp3o” และ “ks-cp3o” ตามลำดับ

2.2 E-divisive

จาก Energy Statistic ให้ $\hat{Q}(X_n, Y_m; \alpha) = \frac{mn}{m+n} \hat{\mathcal{E}}(X_n, Y_m; \alpha)$ คือ ค่าวัดความแตกต่างระหว่างการแจกแจง และ ให้ $Z_1, Z_2, \dots, Z_T \in \mathbb{R}^d$ เป็นลำดับที่อิสระต่อกัน $1 \leq \tau < \kappa \leq T$ เป็นค่าคงที่ และ ให้ $X_\tau = \{Z_1, Z_2, \dots, Z_\tau\}$ และ $Y_\tau(\kappa) = \{Z_{\tau+1}, Z_{\tau+2}, \dots, Z_\kappa\}$ จะได้ว่าจุดเปลี่ยนแปลง $\hat{\tau}$ คือ

$$(\hat{\tau}, \hat{\kappa}) = \underset{(\tau, \kappa)}{\operatorname{argmax}} \hat{Q}(X_\tau, Y_\tau(\kappa); \alpha)$$

สมมติให้มีจุดเปลี่ยนแปลงจำนวน $k - 1$ จุดที่ตำแหน่ง $0 < \hat{\tau}_1 < \dots < \hat{\tau}_{k-1} < T$ จะได้ว่าเป็นการแบ่งข้อมูลออกเป็น k กลุ่ม $\hat{C}_1, \hat{C}_2, \dots, \hat{C}_k$ โดยที่ $\hat{C}_i = \{Z_{\hat{\tau}_{i-1}+1}, \dots, Z_{\hat{\tau}_i+1}\}$ เมื่อ $\hat{\tau}_0 = 0$ และ $\hat{\tau}_k = T$ โดยในแต่ละกลุ่มสามารถใช้ขั้นตอนการหาจุดเปลี่ยนแปลงจุดเดียวภายในกลุ่ม สำหรับในกลุ่มที่ i ให้ $\hat{\tau}(i)$ แทนตำแหน่งที่มีการเปลี่ยนแปลง และ $\hat{\kappa}(i)$ แทนค่าคงที่ที่เกี่ยวข้อง ให้

$$i^* = \underset{i \in \{1, \dots, k\}}{\operatorname{argmax}} \hat{Q}(X_{\hat{\tau}(i)}, Y_{\hat{\tau}(i)}(\hat{\kappa}(i)); \alpha)$$

และตัวสถิติทดสอบ

$$\hat{q}_k = \hat{Q}(X_{\hat{\tau}_k}, Y_{\hat{\tau}_k}(\hat{\kappa}_k); \alpha)$$

โดยที่ $\hat{\tau}_k = \hat{\tau}(i^*)$ คือตำแหน่งของจุดเปลี่ยนที่ k ซึ่งอยู่ภายใน \hat{C}_i และ $\hat{\kappa}_k = \hat{\kappa}(i^*)$ คือค่าคงที่ที่เกี่ยวข้อง

ในการทดสอบนัยสำคัญทางสถิติของจุดเปลี่ยนแปลงนั้นสามารถทำได้โดย

ขั้นตอนที่ 1 ทำการเรียงสับเปลี่ยนข้อมูลภายในแต่ละกลุ่ม

ขั้นตอนที่ 2 ทำตามกระบวนการที่กล่าวมาข้างต้นสำหรับข้อมูลที่ถูกระเรียงสับเปลี่ยนแล้ว

โดยจากการทำขั้นตอนที่ 1 และ 2 จำนวนครั้งจำนวนการเรียงสับเปลี่ยน จะได้ตัวสถิติทดสอบ $\hat{q}_k^{(r)}$ ซึ่งการหาค่า p-value จากการเรียงสับเปลี่ยนทุกวิธีอาจไม่สามารถคำนวณออกมาได้จึงต้องคำนวณค่าประมาณ p-value จากการทำขั้นตอนที่ 1 และ 2 จากจำนวนวิธีการเรียงสับเปลี่ยน R วิธี โดยค่าประมาณ p-value จะเท่ากับ $\#\{r: \hat{q}_k^{(r)} \geq \hat{q}_k\} / R + 1$ โดยหากค่าประมาณ p-value น้อยกว่าค่าระดับนัยสำคัญจะได้ว่า ตำแหน่งที่ k มีการเปลี่ยนแปลง

2.3 Rand index (Rand, 1971)

ให้เซต $S = \{O_1, O_2, \dots, O_n\}$ มีสมาชิก n ตัว โดยแบ่งเซต S เป็น 2 ผลแบ่งกัน (partition) ที่ต่างกัน $U = \{u_1, u_2, \dots, u_R\}$ และ $V = \{v_1, v_2, \dots, v_C\}$ โดยที่ $\bigcup_{i=1}^R u_i = S = \bigcup_{j=1}^C v_j$ และ $u_i \cap u_i = \emptyset = v_j \cap v_j$ สำหรับ $1 \leq i \neq i' \leq R$ และ $1 \leq j \neq j' \leq C$

$$Rand = \frac{a + d}{a + b + c + d}$$

โดยที่ a คือ จำนวนคู่ของสมาชิกที่อยู่ใน U และอยู่ใน V

b คือ จำนวนคู่ของสมาชิกที่อยู่ใน U แต่ไม่อยู่ใน V

c คือ จำนวนคู่ของสมาชิกที่ไม่อยู่ใน U แต่อยู่ใน V

d คือ จำนวนคู่ของสมาชิกที่ไม่อยู่ใน U และไม่อยู่ใน V

โดยค่าของ Rand Index จะอยู่ระหว่าง 0 และ 1 เมื่อผลแบ่งกันทั้ง 2 ผลแบ่งกันเหมือนกัน ค่า Rand Index จะมีค่าเท่ากับ 1 แต่ปัญหาของค่า Rand Index นั้นอยู่ที่ว่าค่าคาดหวังของ Rand Index นั้นไม่เป็นค่าคงที่เมื่อเปรียบเทียบ 2 ผลแบ่งกันที่สุ่มขึ้นมา (random partition) ดังนั้น (Hubert & Arabie, 1985) จึงได้ใช้ตัวแบบไฮเพอร์จีโอเมตริกซึ่งได้มีเงื่อนไขบนทั้งจำนวนกลุ่มที่แบ่งและขนาด และได้นิยาม Adjusted Rand Index ไว้ว่า

$$Adjusted\ Rand = \frac{\sum_{i,j} \binom{n_{ij}}{2} - [\sum_i \binom{n_{i\cdot}}{2} \sum_j \binom{n_{\cdot j}}{2}] / \binom{n}{2}}{\frac{1}{2} [\sum_i \binom{n_{i\cdot}}{2} + \sum_j \binom{n_{\cdot j}}{2}] - [\sum_i \binom{n_{i\cdot}}{2} \sum_j \binom{n_{\cdot j}}{2}] / \binom{n}{2}}$$

โดยที่ $n_{i\cdot}$ คือ จำนวนสมาชิกใน u_i

$n_{\cdot j}$ คือ จำนวนสมาชิกใน v_j

n_{ij} คือ จำนวนสมาชิกที่อยู่ทั้ง u_i และ v_j

2.4 การทดสอบความแตกต่างระหว่างค่าเฉลี่ยสองประชากร (วานิชย์บัญชา, 2540)

เป็นการทดสอบผลต่างระหว่างค่าเฉลี่ยสองลักษณะที่สนใจของ 2 ประชากรว่าแตกต่างกันหรือไม่ โดยใช้ข้อมูลตัวอย่าง 2 ชุด สุ่มจากประชากรทั้งสองอย่างเป็นอิสระกัน ถ้าให้ n_1 และ n_2 เป็นขนาดตัวอย่างที่สุ่มจากประชากรที่ 1 และ 2 ตามลำดับ n_1 และ n_2 จะมีขนาดเท่ากันหรือไม่เท่ากันก็ได้ เมื่อสุ่มตัวอย่างขนาด n_1 จากประชากรที่ 1 และสุ่มตัวอย่างขนาด n_2 จากประชากรที่ 2 อย่างเป็นอิสระกันแล้ว จะนำข้อมูลจากหน่วยตัวอย่างมาทดสอบสมมติฐาน

$$H_0: \mu_1 - \mu_2 = d_0$$

$$H_1: \mu_1 - \mu_2 > d_0 \text{ หรือ } H_1: \mu_1 - \mu_2 < d_0 \text{ หรือ } H_1: \mu_1 - \mu_2 \neq d_0$$

การทดสอบสมมติฐานที่ตั้งขึ้นข้างต้นทั้งแบบข้างเดียวและสองข้างนั้น จะสุ่มตัวอย่างจากแต่ละชุดจากแต่ละประชากรอย่างเป็นอิสระกัน คำนวณค่าเฉลี่ยตัวอย่าง \bar{X}_1 และ \bar{X}_2 ตามลำดับ และจะพิจารณาเป็น 3 กรณีคือ

2.4.1 ประชากรทั้งสองมีการแจกแจงแบบปกติ และทราบค่า σ_1^2, σ_2^2

$$\text{สถิติทดสอบคือ } Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

ซึ่งมีเขตปฏิเสธ H_0 คือ

$$1. Z > Z_{1-\alpha} \text{ เมื่อ } H_1: \mu_1 - \mu_2 > d_0 \text{ ที่ระดับนัยสำคัญ } \alpha$$

$$2. Z < -Z_{1-\alpha} \text{ เมื่อ } H_1: \mu_1 - \mu_2 < d_0 \text{ ที่ระดับนัยสำคัญ } \alpha$$

3. $|Z| > Z_{1-\alpha/2}$ เมื่อ $H_1: \mu_1 - \mu_2 \neq d_0$ ที่ระดับนัยสำคัญ α

2.4.2 ประชากรทั้งสองมีการแจกแจงแบบใด ๆ และขนาดตัวอย่างใหญ่ ($n_1 \geq 30$ และ $n_2 \geq 30$)

1. ทราบค่า σ_1^2 และ σ_2^2

$$\text{สถิติทดสอบคือ } Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

ซึ่งมีเขตปฏิเสธ H_0 คือ

1. $Z > Z_{1-\alpha}$ เมื่อ $H_1: \mu_1 - \mu_2 > d_0$ ที่ระดับนัยสำคัญ α

2. $Z < -Z_{1-\alpha}$ เมื่อ $H_1: \mu_1 - \mu_2 < d_0$ ที่ระดับนัยสำคัญ α

3. $|Z| > Z_{1-\alpha/2}$ เมื่อ $H_1: \mu_1 - \mu_2 \neq d_0$ ที่ระดับนัยสำคัญ α

2.2 ไม่ทราบค่า σ_1^2 และ σ_2^2

$$\text{สถิติทดสอบคือ } Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

ซึ่งมีเขตปฏิเสธ H_0 คือ

1. $Z > Z_{1-\alpha}$ เมื่อ $H_1: \mu_1 - \mu_2 > d_0$ ที่ระดับนัยสำคัญ α

2. $Z < -Z_{1-\alpha}$ เมื่อ $H_1: \mu_1 - \mu_2 < d_0$ ที่ระดับนัยสำคัญ α

3. $|Z| > Z_{1-\alpha/2}$ เมื่อ $H_1: \mu_1 - \mu_2 \neq d_0$ ที่ระดับนัยสำคัญ α

2.4.3 ประชากรทั้งสองมีการแจกแจงแบบปกติ หรือใกล้เคียงแบบปกติ ขนาดตัวอย่างเล็ก ($n_1 < 30$ หรือ $n_2 < 30$) และไม่ทราบค่า σ_1^2 และ σ_2^2

1. ไม่ทราบค่า σ_1^2 และ σ_2^2 ทราบเพียงแต่ว่า $\sigma_1^2 = \sigma_2^2$

สถิติทดสอบคือ $t = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ ที่ระดับองศาอิสระ $n_1 + n_2 - 2$ เมื่อ

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

ซึ่งมีเขตปฏิเสธ H_0 คือ

1. $t > t_{1-\alpha, n_1+n_2-2}$ เมื่อ $H_1: \mu_1 - \mu_2 > d_0$ ที่ระดับนัยสำคัญ

α

2. $t < -t_{1-\alpha, n_1+n_2-2}$ เมื่อ $H_1: \mu_1 - \mu_2 < d_0$ ที่ระดับ

นัยสำคัญ α

3. $|t| > t_{1-\alpha/2, n_1+n_2-2}$ เมื่อ $H_1: \mu_1 - \mu_2 \neq d_0$ ที่ระดับ

นัยสำคัญ α

3.2 ไม่ทราบค่า σ_1^2 และ σ_2^2 แต่ทราบว่า $\sigma_1^2 \neq \sigma_2^2$

สถิติทดสอบคือ $t = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ ที่ระดับองศาอิสระ $v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}}$

ซึ่งมีเขตปฏิเสธ H_0 คือ

1. $t > t_{1-\alpha, v}$ เมื่อ $H_1: \mu_1 - \mu_2 > d_0$ ที่ระดับนัยสำคัญ α

2. $t < -t_{1-\alpha, v}$ เมื่อ $H_1: \mu_1 - \mu_2 < d_0$ ที่ระดับนัยสำคัญ α

3. $|t| > t_{1-\alpha/2, v}$ เมื่อ $H_1: \mu_1 - \mu_2 \neq d_0$ ที่ระดับนัยสำคัญ α

2.5 การทดสอบความแตกต่างระหว่างค่าความแปรปรวนสองประชากร

ให้ S_1^2 และ S_2^2 เป็นค่าความแปรปรวนของตัวอย่างที่เลือกจากประชากรที่ 1 และ 2 ตามลำดับ และประชากรทั้งสองมีการแจกแจงแบบปกติ การทดสอบสมมติฐานเกี่ยวกับความแตกต่างระหว่างค่าความแปรปรวนของสองประชากร

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 > \sigma_2^2 \text{ หรือ } H_1: \sigma_2^2 > \sigma_1^2 \text{ หรือ } H_1: \sigma_1^2 \neq \sigma_2^2$$

2.5.1 การทดสอบสมมติฐานแบบข้างเดียว

$$1. H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 > \sigma_2^2$$

$$\text{สถิติทดสอบคือ } F = \frac{S_1^2}{S_2^2}$$

เขตปฏิเสธ H_0 คือ จะปฏิเสธ H_0 ถ้า $F > F_{1-\alpha}$ ที่ระดับองศาอิสระ $n_1 - 1$ และ $n_2 - 1$

$$2. H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_2^2 > \sigma_1^2$$

$$\text{สถิติทดสอบคือ } F = \frac{S_2^2}{S_1^2}$$

เขตปฏิเสธ H_0 คือ จะปฏิเสธ H_0 ถ้า $F > F_{1-\alpha}$ ที่ระดับองศาอิสระ $n_2 - 1$ และ $n_1 - 1$

2.5.2 การทดสอบสมมติฐานแบบสองข้าง

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

สถิติทดสอบคือ $F = \frac{S_i^2}{S_j^2}$ ถ้า $S_i^2 > S_j^2$ เมื่อ $i = 1, 2$ และ $j = 1, 2$

เขตปฏิเสธคือ จะปฏิเสธ H_0 ถ้า $F > F_{1-\alpha/2}$ ที่ระดับองศาอิสระ $n_i - 1$ และ $n_j - 1$

บทที่ 3

ขอบเขตการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาและเปรียบเทียบประสิทธิภาพวิธีการตรวจหาจุดเปลี่ยนแปลง สำหรับข้อมูลอนุกรมเวลาซึ่งมีขอบเขตการศึกษาดังต่อไปนี้

3.1 เงื่อนไขที่ทำการศึกษา

1. ตัวแปรที่จะศึกษา

ตัวแปรอิสระ (Independent variables ; X) เป็นตัวแปรเชิงปริมาณ 2 มิติ

2. กำหนดขนาดตัวอย่าง 90, 150 และ 300 โดยในแต่ละขนาดตัวอย่างจะแบ่งข้อมูล ออกเป็น 3 กลุ่มซึ่งแต่ละกลุ่มจะมีขนาดตัวอย่างเท่ากัน และรูปแบบการแจกแจงแบบ X, Y, X ตามลำดับ เมื่อ X และ Y มีรูปแบบการแจกแจงที่เหมือนกันแต่ค่าพารามิเตอร์ไม่เท่ากัน

3. กำหนดรูปแบบการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) X และ Y ที่มีค่าเฉลี่ยและค่าความแปรปรวนร่วมต่างกัน โดยที่ μ แทนเวกเตอร์ ค่าเฉลี่ย และ Σ แทนเมทริกซ์ความแปรปรวนร่วม

4. ในการวัดค่าประสิทธิภาพของแต่ละวิธี พิจารณาจากค่า Adjusted Rand Index

5. ในการวิจัยครั้งนี้ ทำการจำลองข้อมูลและทำซ้ำ 500 ครั้ง ในแต่ละกรณีของการทดลอง

3.2 วิธีดำเนินการวิจัย

ส่วนที่ 1 : ศึกษาโดยใช้ข้อมูลจำลอง

1. กำหนดขนาดตัวอย่างเท่ากับ 90, 150 และ 300

2. จำลองข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) โดยกำหนดเวกเตอร์ค่าเฉลี่ยและเมทริกซ์ค่าความแปรปรวนร่วมดังนี้

$$\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

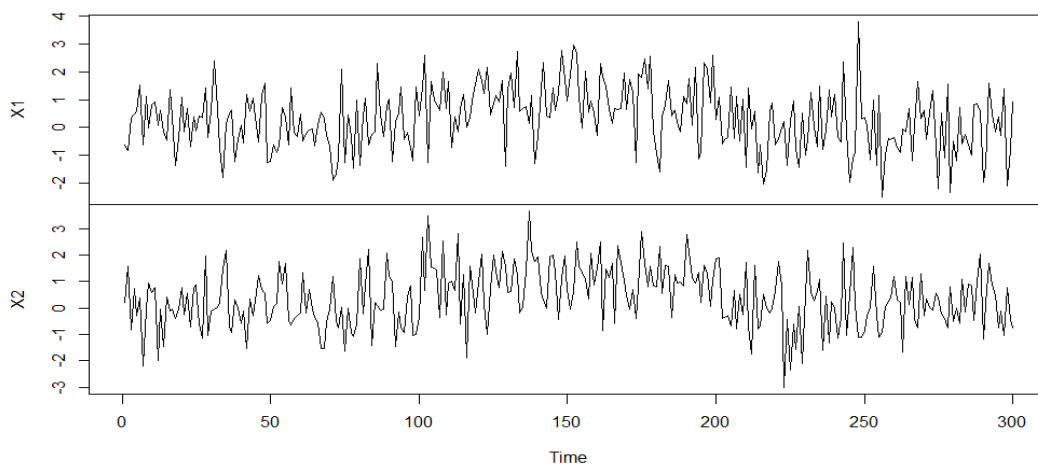
$$\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix},$$

$$\Sigma_3 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \Sigma_4 = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}, \Sigma_5 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

และแบ่งเป็น 8 กรณี ดังนี้

2.1 กรณีที่ 1 ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย $X \sim \mathcal{N}(\mu_1, \Sigma_1)$ และ $Y \sim \mathcal{N}(\mu_2, \Sigma_1)$

Example Case 1

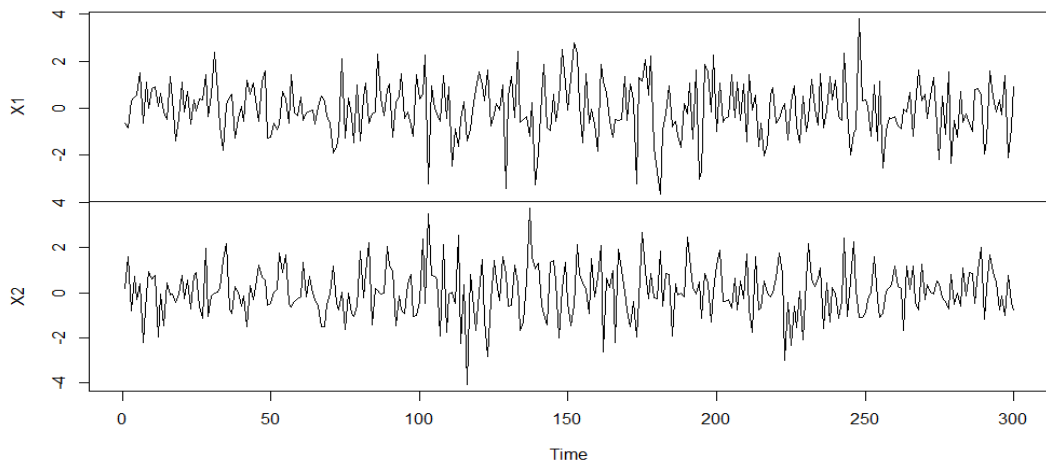


รูปที่ 5 ตัวอย่างข้อมูลที่มีการเปลี่ยนแปลงในค่าเฉลี่ย

โดยตัวอย่างข้อมูลนี้มีการเปลี่ยนแปลงเวกเตอร์เฉลี่ยจาก $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ เป็น $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ในตำแหน่งที่ 101 จนถึง 200 และมีการเปลี่ยนแปลงเวกเตอร์เฉลี่ยจาก $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ เป็น $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ ในตำแหน่งที่ 201 เป็นต้นไป

2.2 กรณีที่ 2 ข้อมูลมีการเปลี่ยนแปลงในค่าความแปรปรวน $X \sim \mathcal{N}(\mu_1, \Sigma_1)$
และ $Y \sim \mathcal{N}(\mu_1, \Sigma_2)$

Example Case 2

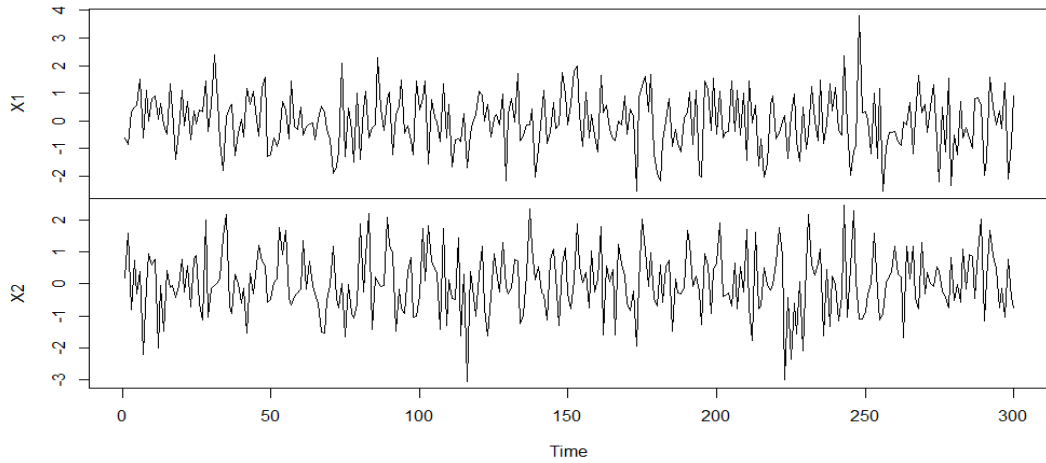


รูปที่ 6 ตัวอย่างข้อมูลที่ข้อมูลมีการเปลี่ยนแปลงในค่าความแปรปรวน

โดยตัวอย่างข้อมูลนี้มีการเปลี่ยนแปลงเมทริกซ์ความแปรปรวนร่วมจาก $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ เป็น $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ ในตำแหน่งที่ 101 จนถึง 200 และมีการเมทริกซ์ความแปรปรวนร่วมจาก $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ เป็น $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ในตำแหน่งที่ 201 เป็นต้นไป

2.3 กรณีที่ 3 ข้อมูลมีการเปลี่ยนแปลงในค่าสหสัมพันธ์ $X \sim \mathcal{N}(\mu_1, \Sigma_1)$ และ $Y \sim \mathcal{N}(\mu_1, \Sigma_3)$

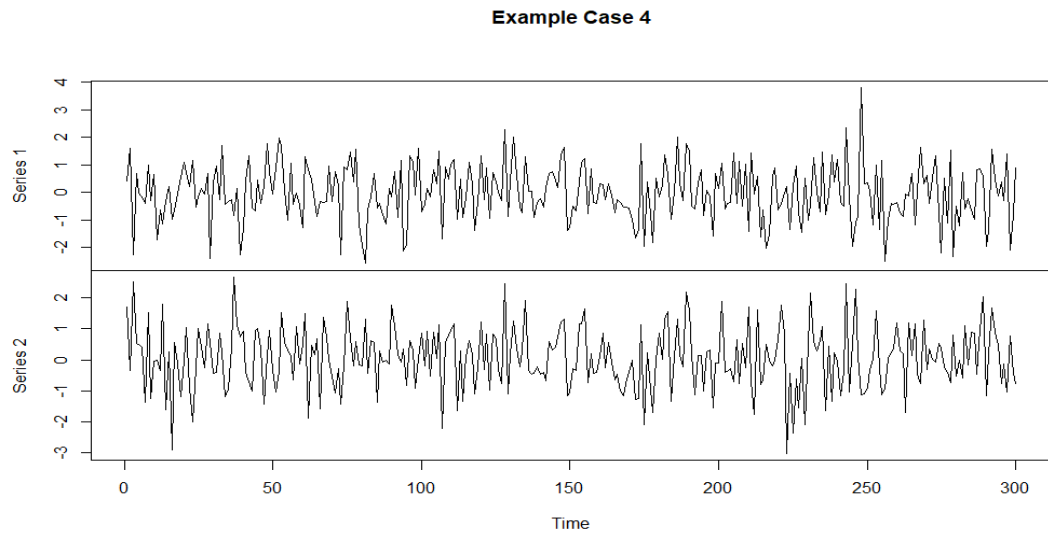
Example Case 3



รูปที่ 7 ตัวอย่างข้อมูลที่ข้อมูลมีการเปลี่ยนแปลงในค่าสหสัมพันธ์

โดยตัวอย่างข้อมูลนี้มีการเปลี่ยนแปลงเมทริกซ์ความแปรปรวนร่วมจาก $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ เป็น $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ ในตำแหน่งที่ 101 จนถึง 200 และมีการเมทริกซ์ความแปรปรวนร่วมจาก $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ เป็น $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ในตำแหน่งที่ 201 เป็นต้นไป

2.4 กรณีที่ 4 ข้อมูลมีการเปลี่ยนแปลงในค่าสหสัมพันธ์ $X \sim \mathcal{N}(\mu_1, \Sigma_1)$ และ $Y \sim \mathcal{N}(\mu_1, \Sigma_4)$



รูปที่ 8 ตัวอย่างข้อมูลที่ข้อมูลมีการเปลี่ยนแปลงในค่าสหสัมพันธ์

โดยตัวอย่างข้อมูลนี้มีการเปลี่ยนแปลงเมทริกซ์ความแปรปรวนร่วมจาก $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ เป็น

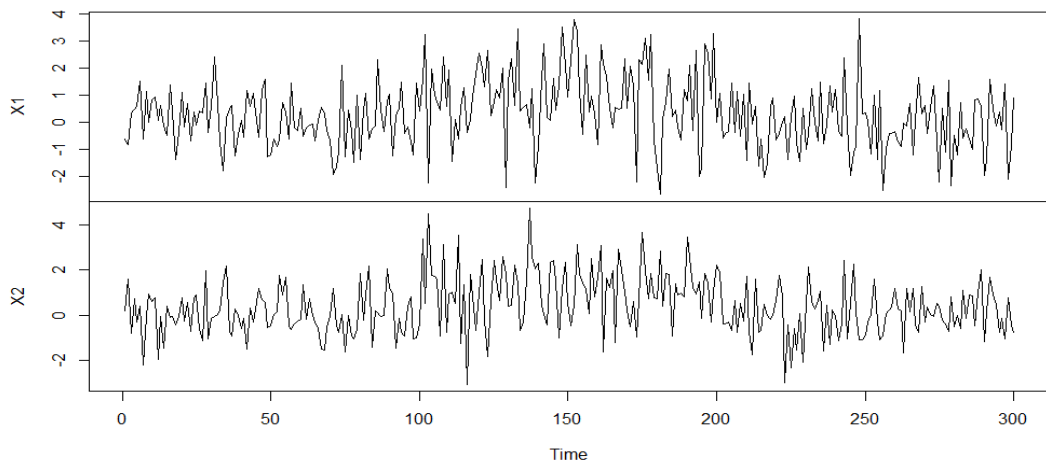
$\begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$ ในตำแหน่งที่ 101 จนถึง 200 และมีการเมทริกซ์ความแปรปรวนร่วมจาก

$\begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$ เป็น $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ในตำแหน่งที่ 201 เป็นต้นไป

2.5 กรณีที่ 5 ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ยและค่าความแปรปรวน

$$X \sim \mathcal{N}(\mu_1, \Sigma_1) \text{ และ } Y \sim \mathcal{N}(\mu_2, \Sigma_2)$$

Example Case 5



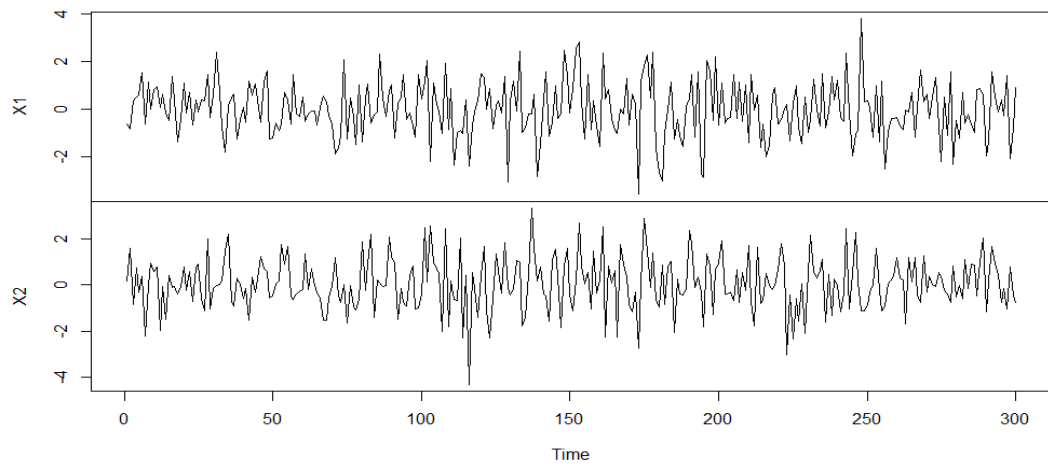
รูปที่ 9 ตัวอย่างข้อมูลที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ยและค่าความแปรปรวน

โดยตัวอย่างข้อมูลนี้มีการเปลี่ยนแปลงเวกเตอร์เฉลี่ยจาก $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ เป็น $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ และเมทริกซ์ความแปรปรวนร่วมจาก $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ เป็น $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ ในตำแหน่งที่ 101 จนถึง 200 และมีการเปลี่ยนแปลงเวกเตอร์เฉลี่ยจาก $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ เป็น $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ และเมทริกซ์ความแปรปรวนร่วมจาก $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ เป็น $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ในตำแหน่งที่ 201 เป็นต้นไป

2.6 กรณีที่ 6 ข้อมูลมีการเปลี่ยนแปลงในค่าความแปรปรวนและค่าสหสัมพันธ์

$$X \sim \mathcal{N}(\mu_1, \Sigma_1) \text{ และ } Y \sim \mathcal{N}(\mu_1, \Sigma_5)$$

Example Case 6



รูปที่ 10 ตัวอย่างข้อมูลที่ข้อมูลมีการเปลี่ยนแปลงในค่าความแปรปรวนและค่าสหสัมพันธ์

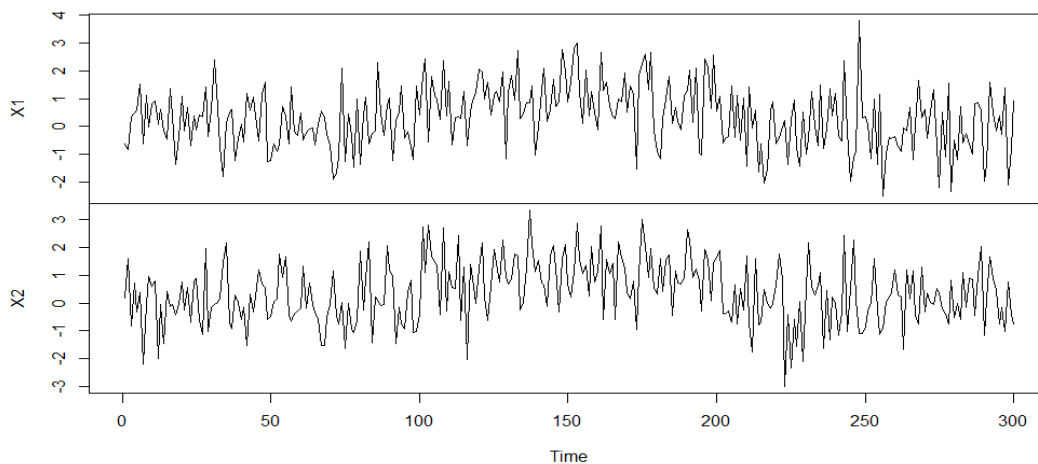
โดยตัวอย่างข้อมูลนี้มีการเปลี่ยนแปลงเมทริกซ์ความแปรปรวนร่วมจาก $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ เป็น

$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ ในตำแหน่งที่ 101 จนถึง 200 และมีการเปลี่ยนแปลงเมทริกซ์ความแปรปรวนร่วมจาก $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ เป็น $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ในตำแหน่งที่ 201 เป็นต้นไป

2.7 กรณีที่ 7 ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ยและค่าสหสัมพันธ์

$$X \sim \mathcal{N}(\mu_1, \Sigma_1) \text{ และ } Y \sim \mathcal{N}(\mu_2, \Sigma_2)$$

Example Case 7

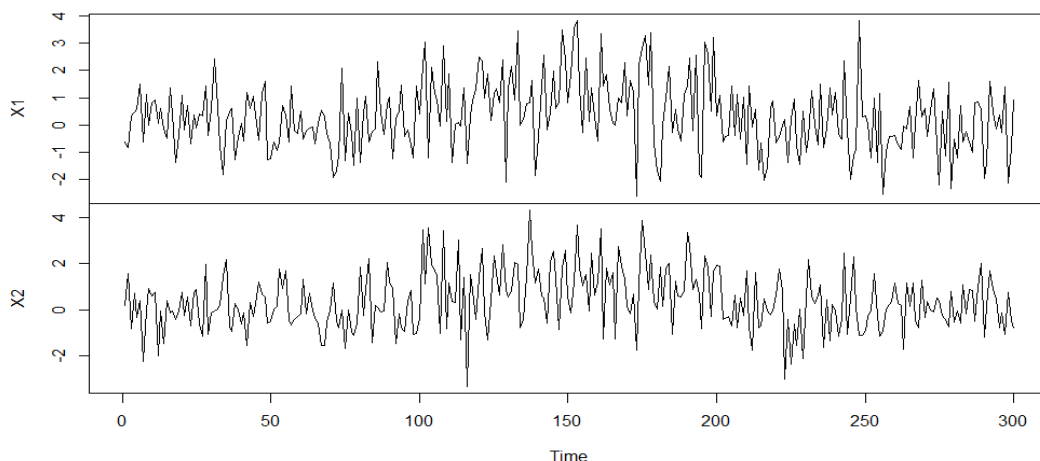


รูปที่ 11 ตัวอย่างข้อมูลที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ยและค่าสหสัมพันธ์

โดยตัวอย่างข้อมูลนี้มีการเปลี่ยนแปลงเวกเตอร์เฉลี่ยจาก $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ เป็น $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ และเมทริกซ์ความแปรปรวนร่วมจาก $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ เป็น $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ ในตำแหน่งที่ 101 จนถึง 200 และมีการเปลี่ยนแปลงเวกเตอร์เฉลี่ยจาก $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ เป็น $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ และเมทริกซ์ความแปรปรวนร่วมจาก $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ เป็น $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ในตำแหน่งที่ 201 เป็นต้นไป

2.8 กรณีที่ 8 ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ $X \sim \mathcal{N}(\mu_1, \Sigma_1)$ และ $Y \sim \mathcal{N}(\mu_2, \Sigma_2)$

Example Case 8



รูปที่ 12 ตัวอย่างข้อมูลที่มีการเปลี่ยนแปลงในค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์

โดยตัวอย่างข้อมูลนี้มีการเปลี่ยนแปลงเวกเตอร์เฉลี่ยจาก $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ เป็น $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ และเมทริกซ์ความแปรปรวนร่วมจาก $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ เป็น $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ ในตำแหน่งที่ 101 จนถึง 200 และมีการเปลี่ยนแปลงเวกเตอร์เฉลี่ยจาก $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ เป็น $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ และเมทริกซ์ความแปรปรวนร่วมจาก $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ เป็น $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ในตำแหน่งที่ 201 เป็นต้นไป

3. ทำการตรวจหาจุดเปลี่ยนแปลงในข้อมูลที่จำลองขึ้นในขั้นตอนที่ 2 โดยใช้วิธี E-Divisive, e-cp3o และ ks-cp3o
4. วัดประสิทธิภาพการตรวจหาจุดเปลี่ยนแปลงตามเกณฑ์ที่ได้ระบุไว้
5. ทำขั้นตอนที่ 1-4 ซ้ำจำนวน 500 รอบในแต่ละกรณี
6. เปรียบเทียบประสิทธิภาพการตรวจหาจุดเปลี่ยนแปลงของแต่ละวิธี
7. สรุปผลที่ได้จากการเปรียบเทียบประสิทธิภาพการตรวจหาจุดเปลี่ยนแปลงว่าวิธีใดตรวจหาจุดเปลี่ยนแปลงได้มีประสิทธิภาพมากที่สุด

ส่วนที่ 2 : ศึกษาโดยใช้ข้อมูลจริง

1. เก็บรวบรวมข้อมูลสัญญาณชีพซึ่งเป็นข้อมูลของมหาวิทยาลัยควีนส์แลนด์ (The university of Queensland) ในกรณีตัวอย่างที่ 2 (Liu, Görges, & Jenkins, 2012) โดยมีตัวแปรที่สนใจคือ อัตราการเต้นของหัวใจ (Heart rate) และอัตราการเต้นของชีพจร (Pulse) ซึ่งมีขนาด 900 วินาที (นาทีที่ 0.12 ถึง นาทีที่ 15.11) และข้อมูลปริมาณฝุ่นละออง PM2.5 จาก BERKELEY EARTH ซึ่งมีขนาด 720 จุด (ข้อมูล ณ วันที่ 1 พฤศจิกายน 2561 ถึง 30 พฤศจิกายน 2561) ซึ่งมีตัวแปรที่สนใจคือ ความหนาแน่นของปริมาณฝุ่นละออง PM2.5 และอุณหภูมิ

2. ตรวจสอบจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o

2.1 ตรวจสอบข้อมูลทั้งหมดในคราวเดียว

2.2 แบ่งข้อมูลออกเป็น 3 ส่วน ส่วนละ 300 วินาที แล้วจึงนำแต่ละส่วนมาตรวจสอบจุดเปลี่ยนแปลง

3. เปรียบเทียบจำนวนและตำแหน่งจุดเปลี่ยนแปลงที่พบใน 2.1 และ 2.2

3.3 ข้อจำกัดในงานวิจัย

1. วิธีการหาจุดเปลี่ยนแปลงที่นำมาเปรียบเทียบกับประสิทธิภาพนั้นเป็นวิธีแบบนอนพาราเมตริกทั้งหมด

2. ข้อมูลที่จำลองขึ้นมา มีรูปแบบการแจกแจงเพียงรูปแบบเดียวคือมีการแจกแจงแบบปกติหลายตัวแปร

3. ข้อมูลที่จำลองขึ้นมา มีจุดเปลี่ยนแปลง 2 จุด

บทที่ 4

ผลการวิจัย

งานวิจัยนี้มีจุดประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการตรวจสอบของวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลอนุกรมเวลาที่มีความผันแปรไม่ปกติที่มีการแจกแจงปกติ 2 ตัวแปร มีการเปลี่ยนแปลงในค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ โดยทำการเปรียบเทียบในส่วนของข้อมูลจำลองที่มีขนาดตัวอย่าง 90, 150 และ 300 โดยจำลองข้อมูลแต่ละขนาดตัวอย่างออกเป็น 8 กรณี

กรณีที่ 1 ข้อมูลมีการเปลี่ยนแปลงค่าเฉลี่ย

กรณีที่ 2 ข้อมูลมีการเปลี่ยนแปลงค่าความแปรปรวน

กรณีที่ 3 ข้อมูลมีการเปลี่ยนแปลงค่าสหสัมพันธ์ระดับปานกลาง ($\rho = 0.5$)

กรณีที่ 4 ข้อมูลมีการเปลี่ยนแปลงค่าสหสัมพันธ์ระดับสูง ($\rho = 0.9$)

กรณีที่ 5 ข้อมูลมีการเปลี่ยนแปลงค่าเฉลี่ยและค่าความแปรปรวน

กรณีที่ 6 ข้อมูลมีการเปลี่ยนแปลงค่าความแปรปรวนและค่าสหสัมพันธ์

กรณีที่ 7 ข้อมูลมีการเปลี่ยนแปลงค่าเฉลี่ยและค่าสหสัมพันธ์

กรณีที่ 8 ข้อมูลมีการเปลี่ยนแปลงค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์

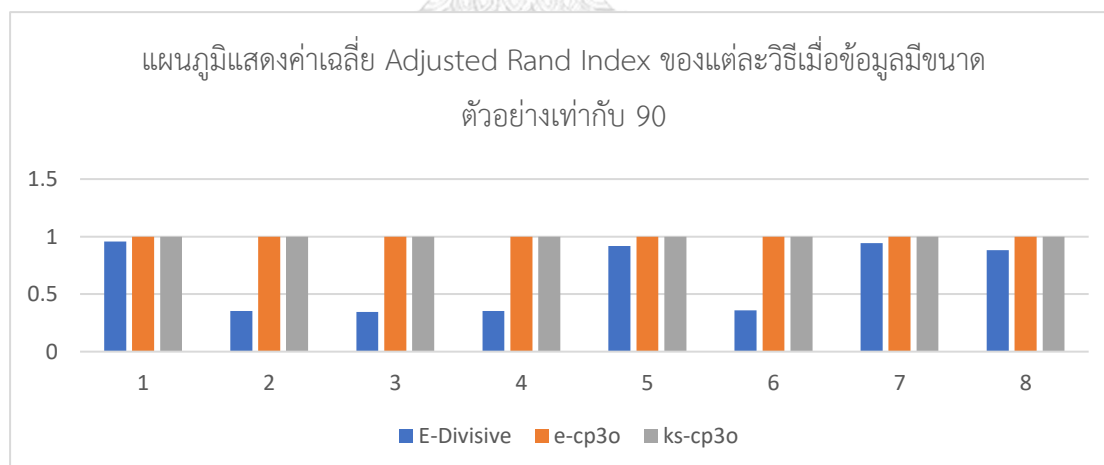
และเปรียบเทียบประสิทธิภาพของแต่ละวิธีโดยพิจารณาค่า Adjusted Rand Index และจำนวนจุดเปลี่ยนแปลงที่พบในแต่ละกรณี และทำการเปรียบเทียบในส่วนของข้อมูลจริง ซึ่งเป็นข้อมูลสัญญาณชีพของมหาวิทยาลัยควีนส์แลนด์ (The university of Queensland) โดยมีตัวแปรที่สนใจคือ อัตราการเต้นของหัวใจ (Heart rate) และอัตราการเต้นของชีพจร (Pulse) ซึ่งมีขนาด 900 วินาที (นาทีที่ 0.12 ถึง นาทีที่ 15.11) และข้อมูลปริมาณฝุ่นละออง PM2.5 จาก BERKELEY EARTH ซึ่งมีขนาด 720 จุด (ข้อมูล ณ วันที่ 1 พฤศจิกายน 2561 ถึง 30 พฤศจิกายน 2561) ซึ่งมีตัวแปรที่สนใจคือ ความหนาแน่นของปริมาณฝุ่นละออง PM2.5 และอุณหภูมิ โดยทำการตรวจหาข้อมูลทั้งหมดในคราวเดียว และแบ่งข้อมูลออกเป็น 3 ส่วนก่อนแล้วจึงนำแต่ละส่วนมาตรวจหาจุดเปลี่ยนแปลง และทำการทดสอบสมมติฐานการเปลี่ยนแปลงของค่าเฉลี่ย และค่าความแปรปรวนในแต่ละช่วงข้อมูล

4.1 ผลการเปรียบเทียบวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o ในส่วนข้อมูลจำลอง

4.1.1 ผลการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 90

ข้อมูลมีการเปลี่ยนแปลงใน	E-Divisive	e-cp3o	ks-cp3o
ค่าเฉลี่ย	0.9563	1.0000	1.0000
ค่าความแปรปรวน	0.3523	1.0000	1.0000
ค่าสหสัมพันธ์ ($\rho = 0.5$)	0.3449	1.0000	1.0000
ค่าสหสัมพันธ์ ($\rho = 0.9$)	0.3523	1.0000	1.0000
ค่าเฉลี่ย และค่าความแปรปรวน	0.9192	1.0000	1.0000
ค่าความแปรปรวน และค่าสหสัมพันธ์	0.3578	1.0000	1.0000
ค่าเฉลี่ย และค่าสหสัมพันธ์	0.9437	1.0000	1.0000
ค่าเฉลี่ย, ค่าความแปรปรวน และค่าสหสัมพันธ์	0.8823	1.0000	1.0000

ตารางที่ 1 ค่าเฉลี่ย Adjusted Rand Index ของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่างเท่ากับ 90



รูปที่ 13 แผนภูมิแสดงค่าเฉลี่ย Adjusted Rand Index ของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่างเท่ากับ 90

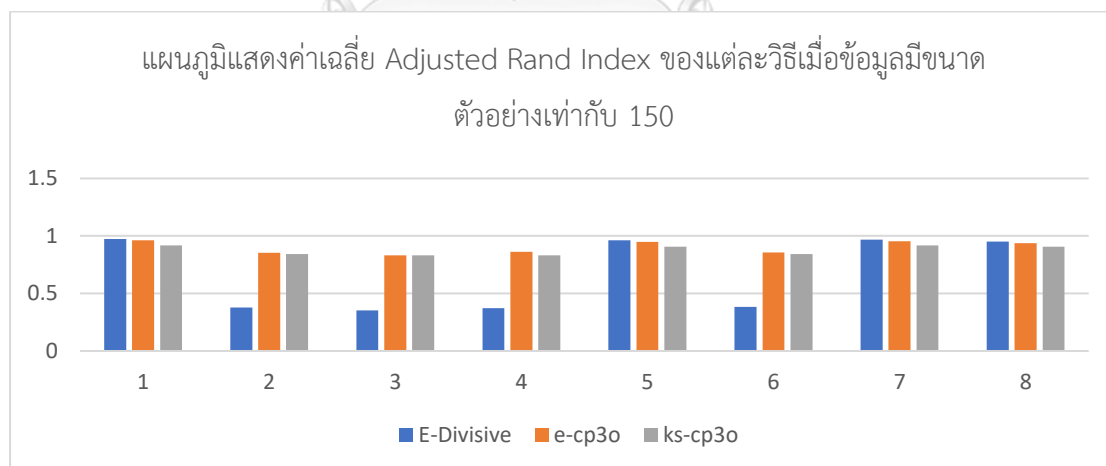
สรุปได้ว่า ในการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 90 พบว่า การตรวจหาจุดเปลี่ยนแปลงด้วยวิธี e-cp3o และ ks-cp3o มีค่า Adjusted Rand Index สูงที่สุดในทุกกรณี การตรวจหาจุด

เปลี่ยนแปลงด้วยวิธี E-Divisive มีค่า Adjusted Rand Index สูงในกรณีที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย และมีค่าต่ำในกรณีอื่น ๆ

4.1.2 ผลการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 150

ข้อมูลมีการเปลี่ยนแปลงใน	E-Divisive	e-cp3o	ks-cp3o
ค่าเฉลี่ย	0.9738	0.9636	0.9190
ค่าความแปรปรวน	0.3788	0.8533	0.8437
ค่าสหสัมพันธ์ ($\rho = 0.5$)	0.3517	0.8313	0.8311
ค่าสหสัมพันธ์ ($\rho = 0.9$)	0.3714	0.8612	0.8325
ค่าเฉลี่ย และค่าความแปรปรวน	0.9632	0.9488	0.9061
ค่าความแปรปรวน และค่าสหสัมพันธ์	0.3832	0.8568	0.8431
ค่าเฉลี่ย และค่าสหสัมพันธ์	0.9673	0.9548	0.9188
ค่าเฉลี่ย, ค่าความแปรปรวน และค่าสหสัมพันธ์	0.9499	0.9382	0.9064

ตารางที่ 2 ค่าเฉลี่ย Adjusted Rand Index ของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่างเท่ากับ 150



รูปที่ 14 แผนภูมิแสดงค่าเฉลี่ย Adjusted Rand Index ของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่างเท่ากับ 150

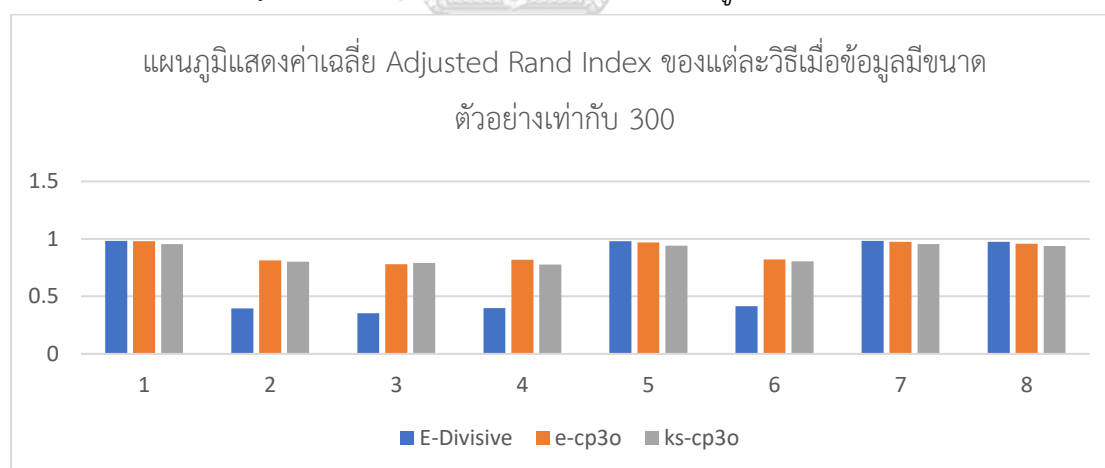
สรุปได้ว่า ในการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 150 พบว่า การตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive มีค่า Adjusted Rand Index สูงที่สุดในกรณีที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย

วิธี e-cp3o มีค่า Adjusted Rand Index สูงที่สุดในกรณีอื่น ๆ วิธี ks-cp3o มีค่า Adjusted Rand Index ที่สูงในทุกกรณี

4.1.3 ผลการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 300

ข้อมูลมีการเปลี่ยนแปลงใน	E-Divisive	e-cp3o	ks-cp3o
ค่าเฉลี่ย	0.9836	0.9797	0.9561
ค่าความแปรปรวน	0.3947	0.8130	0.8020
ค่าสหสัมพันธ์ ($\rho = 0.5$)	0.3531	0.7786	0.7914
ค่าสหสัมพันธ์ ($\rho = 0.9$)	0.3972	0.8187	0.7779
ค่าเฉลี่ย และค่าความแปรปรวน	0.9791	0.9704	0.9399
ค่าความแปรปรวน และค่าสหสัมพันธ์	0.4159	0.8224	0.8034
ค่าเฉลี่ย และค่าสหสัมพันธ์	0.9818	0.9732	0.9557
ค่าเฉลี่ย, ค่าความแปรปรวน และค่าสหสัมพันธ์	0.9738	0.9571	0.9377

ตารางที่ 3 ค่าเฉลี่ย Adjusted Rand Index ของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่างเท่ากับ 300



รูปที่ 15 แผนภูมิแสดงค่าเฉลี่ย Adjusted Rand Index ของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่างเท่ากับ 300

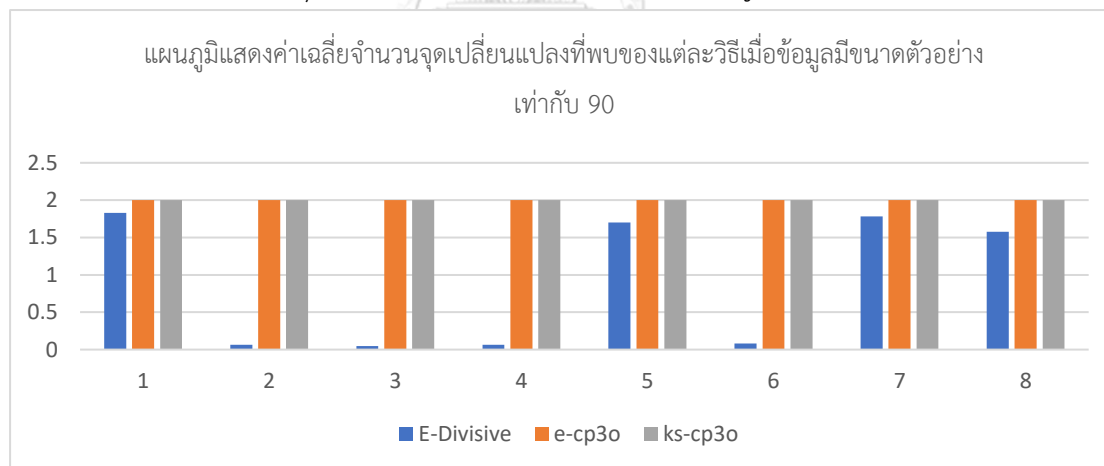
สรุปได้ว่า ในการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 300 พบว่า การตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive มีค่า Adjusted Rand Index สูงที่สุดในกรณีที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย

วิธี e-cp3o มีค่า Adjusted Rand Index สูงที่สุดในกรณีอื่น ๆ ยกเว้นกรณีที่ข้อมูลมีการเปลี่ยนแปลงค่าสหสัมพันธ์ ($\rho = 0.5$) วิธี ks-cp3o มีค่า Adjusted Rand Index ที่สูงที่สุด

4.1.4 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 90

ข้อมูลมีการเปลี่ยนแปลงใน	E-Divisive	e-cp3o	ks-cp3o
ค่าเฉลี่ย	1.8300	2.0000	2.0000
ค่าความแปรปรวน	0.0660	2.0000	2.0000
ค่าสหสัมพันธ์ ($\rho = 0.5$)	0.0460	2.0000	2.0000
ค่าสหสัมพันธ์ ($\rho = 0.9$)	0.0640	2.0000	2.0000
ค่าเฉลี่ย และค่าความแปรปรวน	1.7000	2.0000	2.0000
ค่าความแปรปรวน และค่าสหสัมพันธ์	0.0800	2.0000	2.0000
ค่าเฉลี่ย และค่าสหสัมพันธ์	1.7840	2.0000	2.0000
ค่าเฉลี่ย, ค่าความแปรปรวน และค่าสหสัมพันธ์	1.5760	2.0000	2.0000

ตารางที่ 4 ค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงที่พบของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่าง 90



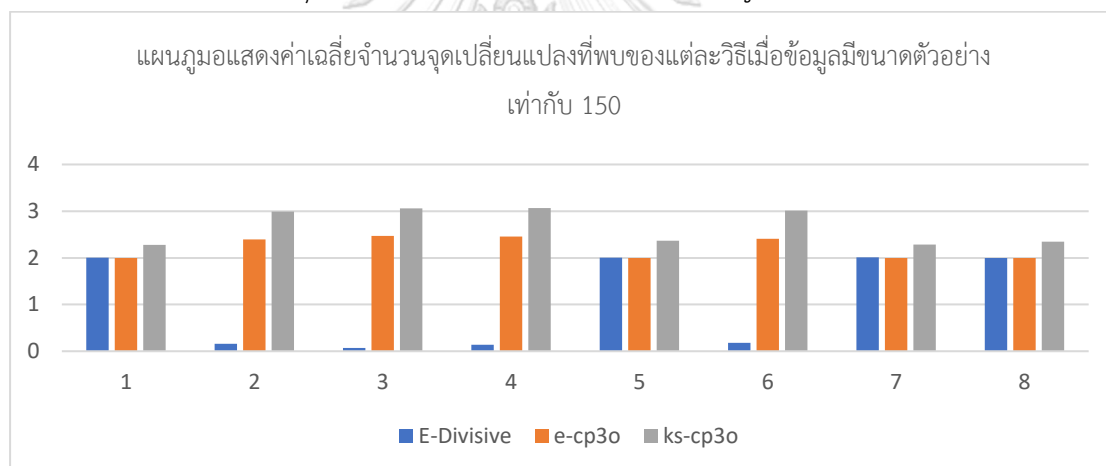
รูปที่ 16 แผนภูมิแสดงค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงที่พบของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่าง 90

สรุปได้ว่า ค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงที่ตรวจพบของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 90 พบว่าโดยเฉลี่ยแล้ววิธี e-cp3o และ ks-cp3o มีการตรวจพบจำนวนจุดที่มีการเปลี่ยนแปลงแม่นยำมากที่สุด วิธี E-Divisive มีการตรวจพบจำนวนจุดที่มีการเปลี่ยนแปลงค่อนข้างแม่นยำในกรณีที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย แต่ไม่พบจุดที่มีการเปลี่ยนแปลงในกรณีอื่น ๆ

4.1.5 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 150

ข้อมูลมีการเปลี่ยนแปลงใน	E-Divisive	e-cp3o	ks-cp3o
ค่าเฉลี่ย	2.0040	2.0000	2.2760
ค่าความแปรปรวน	0.1600	2.3960	2.9940
ค่าสหสัมพันธ์ ($\rho = 0.5$)	0.0720	2.4700	3.0620
ค่าสหสัมพันธ์ ($\rho = 0.9$)	0.1380	2.4540	3.0640
ค่าเฉลี่ย และค่าความแปรปรวน	2.0060	2.0000	2.3660
ค่าความแปรปรวน และค่าสหสัมพันธ์	0.1800	2.4080	3.0100
ค่าเฉลี่ย และค่าสหสัมพันธ์	2.0120	2.000	2.2860
ค่าเฉลี่ย, ค่าความแปรปรวน และค่าสหสัมพันธ์	1.9940	2.000	2.3480

ตารางที่ 5 ค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงที่พบของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่าง 150



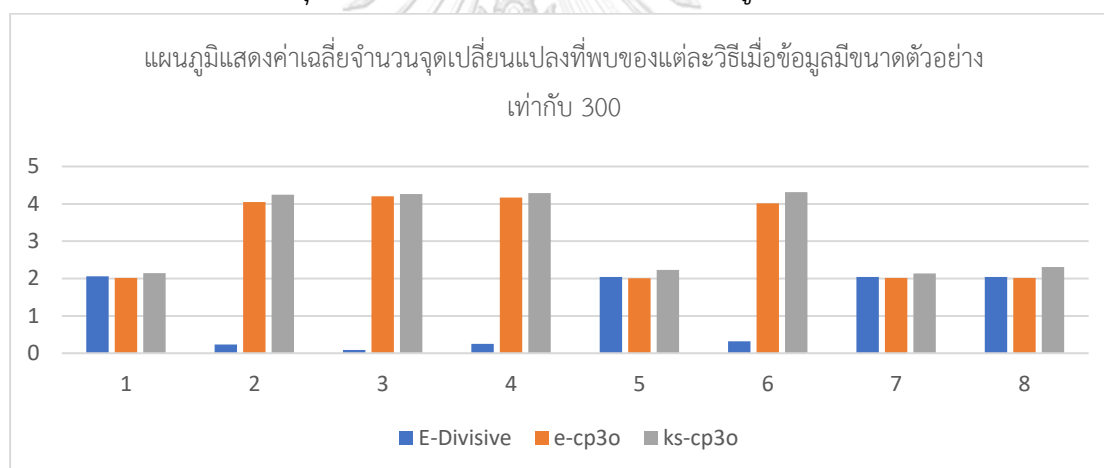
รูปที่ 17 แผนภูมิแสดงค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงที่พบของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่าง 150

สรุปได้ว่า ค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงที่ตรวจพบของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 150 พบว่า โดยเฉลี่ยแล้ววิธี e-cp3o มีการตรวจพบจำนวนจุดที่แม่นยำมากที่สุด วิธี E-Divisive มีการตรวจพบจำนวนจุดที่แม่นยำเฉพาะกรณีที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย แต่ไม่พบจุดที่มีการเปลี่ยนแปลงเมื่อข้อมูลมีการเปลี่ยนแปลงในกรณีอื่น ๆ วิธี ks-cp3o มีการตรวจพบเจอจำนวนจุดโดยเฉลี่ยมากที่สุด 3 วิธี

4.1.6 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 300

ข้อมูลมีการเปลี่ยนแปลงใน	E-Divisive	e-cp3o	ks-cp3o
ค่าเฉลี่ย	2.0580	2.0120	2.1480
ค่าความแปรปรวน	0.2300	4.0460	4.2440
ค่าสหสัมพันธ์ ($\rho = 0.5$)	0.0880	4.2000	4.2640
ค่าสหสัมพันธ์ ($\rho = 0.9$)	0.2500	4.1660	4.2880
ค่าเฉลี่ย และค่าความแปรปรวน	2.0420	2.0100	2.2300
ค่าความแปรปรวน และค่าสหสัมพันธ์	0.3200	4.0140	4.3120
ค่าเฉลี่ย และค่าสหสัมพันธ์	2.0400	2.0120	2.1360
ค่าเฉลี่ย, ค่าความแปรปรวน และค่าสหสัมพันธ์	2.0380	2.0200	2.3040

ตารางที่ 6 ค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงที่พบของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่าง 300



รูปที่ 18 แผนภูมิแสดงค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงที่พบของแต่ละวิธีเมื่อข้อมูลมีขนาดตัวอย่าง 300

สรุปได้ว่า ค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงที่ตรวจพบของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 300 พบว่า โดยเฉลี่ยแล้ววิธี E-Divisive, e-cp3o และ ks-cp3o มีการตรวจพบจำนวนจุดที่แม่นยำในกรณีที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย แต่ในกรณีที่ข้อมูลมีการเปลี่ยนแปลงกรณีอื่น ๆ โดยเฉลี่ยวิธี E-Divisive ไม่พบจุดที่มีการเปลี่ยนแปลง วิธี e-cp3o และ ks-cp3o มีการตรวจเจอจุดที่มีการเปลี่ยนแปลงไม่แม่นยำนัก

4.1.7 ผลการเปรียบเทียบประสิทธิภาพและจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive และ e-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 90

ข้อมูลที่มีการเปลี่ยนแปลงใน	E-Divisive		e-cp3o	
	ค่าเฉลี่ย Adjusted Rand Index	ค่าเฉลี่ย จำนวนจุดที่พบ	ค่าเฉลี่ย Adjusted Rand Index	ค่าเฉลี่ย จำนวนจุดที่พบ
ค่าเฉลี่ย	0.9563	1.8300	1.0000	2.0000
ค่าความแปรปรวน	0.3523	0.0660	1.0000	2.0000
ค่าสหสัมพันธ์ ($\rho = 0.5$)	0.3449	0.0460	1.0000	2.0000
ค่าสหสัมพันธ์ ($\rho = 0.9$)	0.3523	0.0640	1.0000	2.0000
ค่าเฉลี่ย และค่าความแปรปรวน	0.9192	1.7000	1.0000	2.0000
ค่าความแปรปรวน และค่าสหสัมพันธ์	0.3578	0.0800	1.0000	2.0000
ค่าเฉลี่ย และค่าสหสัมพันธ์	0.9437	1.7840	1.0000	2.0000
ค่าเฉลี่ย, ค่าความแปรปรวน และค่าสหสัมพันธ์	0.8823	1.5760	1.0000	2.0000

ตารางที่ 7 ค่าเฉลี่ย Adjusted Rand Index และค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงของวิธี E-Divisive และ e-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 90

สรุปได้ว่า เมื่อเปรียบเทียบค่าเฉลี่ย Adjusted Rand Index และค่าเฉลี่ยจำนวนจุดที่มีการเปลี่ยนแปลงที่พบของวิธี E-Divisive และ e-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 90 พบว่า ค่าเฉลี่ย Adjusted Rand Index และค่าเฉลี่ยจำนวนจุดที่มีการเปลี่ยนแปลงที่พบของวิธี e-cp3o มีค่าสูงกว่า และใกล้เคียงมากกว่าวิธี E-Divisive ทั้งสองค่า

4.1.8 ผลการเปรียบเทียบประสิทธิภาพและจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive และ e-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 150

ข้อมูลมีการเปลี่ยนแปลงใน	E-Divisive		e-cp3o	
	ค่าเฉลี่ย Adjusted Rand Index	ค่าเฉลี่ย จำนวนจุดที่พบ	ค่าเฉลี่ย Adjusted Rand Index	ค่าเฉลี่ย จำนวนจุดที่พบ
ค่าเฉลี่ย	0.9738	2.0040	0.9636	2.0000
ค่าความแปรปรวน	0.3788	0.1600	0.8533	2.3960
ค่าสหสัมพันธ์ ($\rho = 0.5$)	0.3517	0.0720	0.8313	2.4700
ค่าสหสัมพันธ์ ($\rho = 0.9$)	0.3714	0.1380	0.8612	2.4540
ค่าเฉลี่ย และค่าความแปรปรวน	0.9632	2.0060	0.9488	2.0000
ค่าความแปรปรวน และค่าสหสัมพันธ์	0.3832	0.1800	0.8568	2.4080
ค่าเฉลี่ย และค่าสหสัมพันธ์	0.9673	2.0120	0.9548	2.0000
ค่าเฉลี่ย, ค่าความแปรปรวน และค่าสหสัมพันธ์	0.9499	1.9940	0.9382	2.0000

ตารางที่ 8 ค่าเฉลี่ย Adjusted Rand Index และค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงของวิธี E-Divisive และ e-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 150

สรุปได้ว่า เมื่อเปรียบเทียบค่าเฉลี่ย Adjusted Rand Index และค่าเฉลี่ยจำนวนจุดที่มีการเปลี่ยนแปลงที่พบของวิธี E-Divisive และ e-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 150 พบว่าในกรณีที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย ค่าเฉลี่ย Adjusted Rand Index ของวิธี E-Divisive มีค่ามากกว่าของวิธี e-cp3o แต่เมื่อเปรียบเทียบค่าเฉลี่ยจำนวนจุดที่มีการเปลี่ยนแปลงที่พบ ค่าเฉลี่ยจำนวนจุดที่มีการเปลี่ยนแปลงที่พบของวิธี e-cp3o นั้นมีความแม่นยำมากกว่าวิธี E-Divisive และในกรณีที่ข้อมูลไม่มีการเปลี่ยนแปลงในค่าเฉลี่ย ค่าเฉลี่ย Adjusted Rand Index และค่าเฉลี่ยจำนวนจุดที่มีการเปลี่ยนแปลงที่พบของวิธี e-cp3o มีค่าสูงกว่าและใกล้เคียงมากกว่าวิธี E-Divisive ทั้งสองค่า

4.1.9 ผลการเปรียบเทียบประสิทธิภาพและจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive และ e-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 300

ข้อมูลมีการเปลี่ยนแปลงใน	E-Divisive		e-cp3o	
	ค่าเฉลี่ย Adjusted Rand Index	ค่าเฉลี่ย จำนวนจุดที่พบ	ค่าเฉลี่ย Adjusted Rand Index	ค่าเฉลี่ย จำนวนจุดที่พบ
ค่าเฉลี่ย	0.9836	2.0580	0.9797	2.0120
ค่าความแปรปรวน	0.3947	0.2300	0.8130	4.0460
ค่าสหสัมพันธ์ ($\rho = 0.5$)	0.3531	0.0880	0.7786	4.2000
ค่าสหสัมพันธ์ ($\rho = 0.9$)	0.3972	0.2500	0.8187	4.1660
ค่าเฉลี่ย และค่าความแปรปรวน	0.9791	2.0420	0.9704	2.0100
ค่าความแปรปรวน และค่าสหสัมพันธ์	0.4159	0.3200	0.8224	4.0140
ค่าเฉลี่ย และค่าสหสัมพันธ์	0.9818	2.0400	0.9732	2.0120
ค่าเฉลี่ย, ค่าความแปรปรวน และค่าสหสัมพันธ์	0.9738	2.0380	0.9571	2.0200

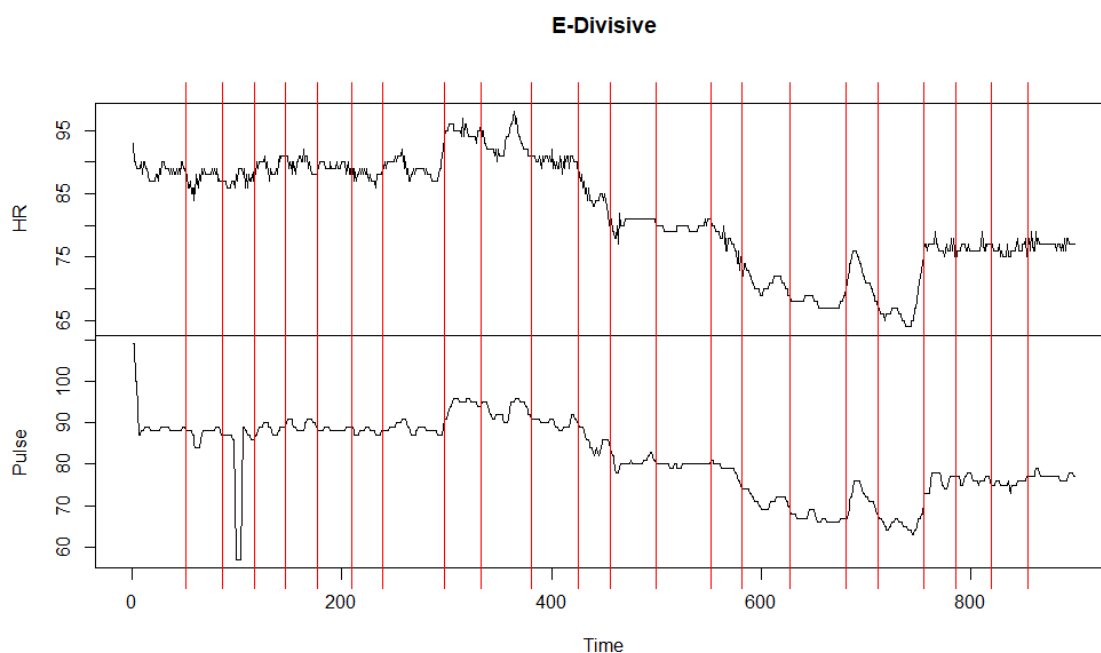
ตารางที่ 9 ค่าเฉลี่ย Adjusted Rand Index และค่าเฉลี่ยจำนวนจุดเปลี่ยนแปลงของวิธี E-Divisive และ e-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 300

สรุปได้ว่า เมื่อเปรียบเทียบค่าเฉลี่ย Adjusted Rand Index และค่าเฉลี่ยจำนวนจุดที่มีการเปลี่ยนแปลงที่พบของวิธี E-Divisive และ e-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 300 พบว่าในกรณีที่มีข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย ค่าเฉลี่ย Adjusted Rand Index ของวิธี E-Divisive มีค่ามากกว่าของวิธี e-cp3o แต่เมื่อเปรียบเทียบค่าเฉลี่ยจำนวนจุดที่มีการเปลี่ยนแปลงที่พบ ค่าเฉลี่ยจำนวนจุดที่มีการเปลี่ยนแปลงที่พบของวิธี e-cp3o นั้นมีความใกล้เคียงมากกว่าวิธี E-Divisive และในกรณีที่ข้อมูลไม่มีการเปลี่ยนแปลงในค่าเฉลี่ย ค่าเฉลี่ย Adjusted Rand Index ของวิธี e-cp3o มีค่าสูงกว่าวิธี E-Divisive แต่ค่าเฉลี่ยจำนวนจุดที่มีการเปลี่ยนแปลงที่พบไม่สามารถสรุปได้ว่าวิธีใดมีความแม่นยำมากกว่ากัน

4.2 ผลการเปรียบเทียบวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o ในส่วนข้อมูลจริง

ในส่วนของข้อมูลจริงนั้นเป็นข้อมูลสัญญาณชีพของมหาวิทยาลัยควีนส์แลนด์ (The university of Queensland) โดยมีตัวแปรที่สนใจคือ อัตราการเต้นของหัวใจ (Heart rate) และ อัตราการเต้นของชีพจร (Pulse) ซึ่งมีขนาด 900 วินาที (นาทีที่ 0.12 ถึง นาทีที่ 15.11) โดยทำการตรวจหาข้อมูลทั้งหมดในคราวเดียว และแบ่งข้อมูลออกเป็น 3 ส่วนก่อนแล้วจึงนำแต่ละส่วนมาตรวจหาจุดเปลี่ยนแปลงซึ่งได้ผลการวิจัย ดังนี้

4.2.1 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงของข้อมูลสัญญาณชีพที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o โดยทำการตรวจหาข้อมูลทั้งหมดในคราวเดียว



รูปที่ 19 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive โดยตรวจข้อมูลทั้งหมด

ผลจากการนำวิธี E-Divisive ไปใช้ในการตรวจหาจุดเปลี่ยนแปลงในข้อมูลสัญญาณชีพ พบว่าวิธี E-Divisive ตรวจพบจุดที่ข้อมูลมีการเปลี่ยนแปลงจำนวน 22 จุด ที่ตำแหน่ง 51, 86, 116, 146, 176, 209, 239, 298, 333, 381, 426, 456, 500, 552, 582, 628, 681, 712, 756, 786, 820 และ 855 ซึ่งมีค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลง ดังนี้

วินาทีที่	ค่าเฉลี่ยอัตรา การเต้นของ หัวใจ	ค่าเฉลี่ยชีพ จร	ค่าความ แปรปรวนของ อัตราการเต้น ของหัวใจ	ค่าความ แปรปรวนของ ชีพจร	ค่า สหสัมพันธ์
1-50	88.78	89.72	1.31	21.77	0.52
51-85	87.37	87.31	1.53	2.52	0.33
86-115	87.27	81.13	1.10	134.74	-0.48
116-145	89.40	88.40	1.42	1.35	0.38
146-175	89.70	89.67	1.46	1.40	-0.27
176-208	89.30	88.33	0.41	0.23	-0.44
209-238	87.87	87.87	0.60	0.60	-0.08
239-297	89.10	88.56	1.75	1.35	0.50
298-332	94.94	94.63	0.76	2.12	-0.02
333-380	93.06	93.25	4.06	3.64	0.58
381-425	90.20	90.07	0.71	1.20	0.38
426-455	84.73	85.53	3.44	4.88	0.68
456-499	80.52	80.59	1.00	1.83	0.28
500-551	79.67	79.83	0.38	0.15	0.09
552-581	77.63	78.87	3.83	2.33	0.86
582-627	71.04	71.37	1.95	2.59	0.85
628-680	67.83	67.08	0.61	0.99	0.74
681-711	72.19	72.03	6.43	7.77	0.75
712-755	66.73	65.66	6.81	2.51	0.74
756-785	76.60	75.90	0.66	3.82	0.26
786-819	76.38	76.47	0.61	0.74	-0.14
820-854	76.09	75.37	0.73	0.59	0.31
855-900	77.04	77.11	0.44	0.59	0.25

ตารางที่ 10 ค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลงเมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive

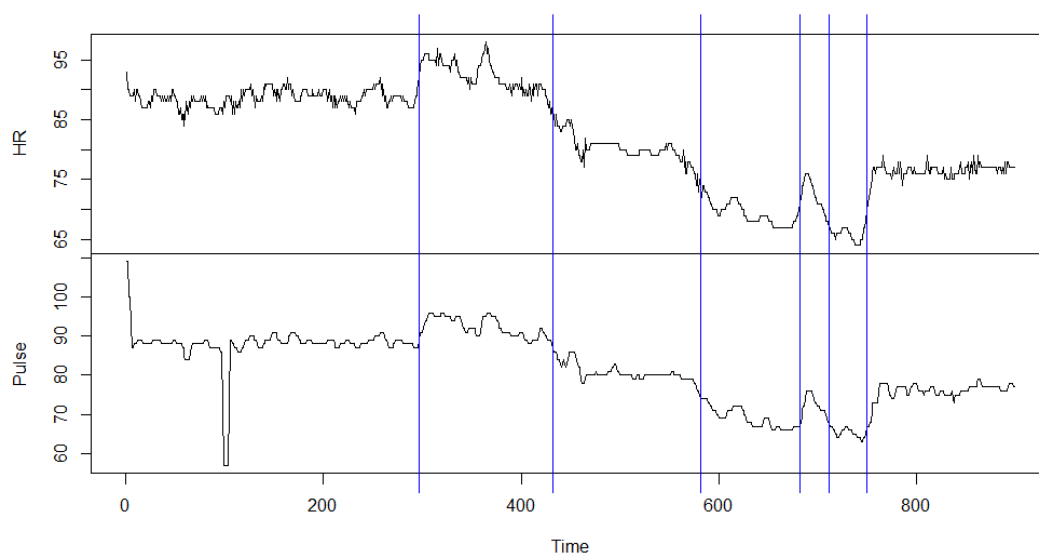
และจากการนำข้อมูลในแต่ละช่วงไปทดสอบสมมติฐานความแตกต่างของค่าเฉลี่ย และค่าความแปรปรวนของสองประชากรโดยใช้สถิติทดสอบ t และ F ตามลำดับ พบว่า ในทุก ๆ ช่วงการเปรียบเทียบ ไม่ว่าจะเป็นการทดสอบความแตกต่างของค่าเฉลี่ยของอัตราการเต้นของหัวใจ ค่าเฉลี่ยของชีพจร ค่าความแปรปรวนของอัตราการเต้นของหัวใจ หรือค่าความแปรปรวนของชีพจร จะมีอย่างน้อยหนึ่งค่าที่มีค่า p-value น้อยกว่า 0.05 หมายถึงค่าระหว่างสองช่วงที่เปรียบเทียบกันนั้นมีความแตกต่างกันทางสถิติที่ระดับนัยสำคัญ 0.05 ซึ่งสอดคล้องกับตำแหน่งจุดเปลี่ยนแปลงที่พบด้วยวิธี E-Divisive

เปรียบเทียบช่วงวินาทีที่	ค่า P-Value			
	ค่าเฉลี่ยอัตราการเต้นของหัวใจ	ค่าเฉลี่ยชีพจร	ค่าความแปรปรวนของอัตราการเต้นของหัวใจ	ค่าความแปรปรวนของชีพจร
1-50 กับ 51-85	0.0000*	0.0012*	0.6171	0.0000*
51-85 กับ 86-115	0.7167	0.0070*	0.3623	0.0000*
86-115 กับ 116-145	0.0000*	0.0019*	0.4935	0.0000*
116-145 กับ 146-175	0.3369	0.0001*	0.9439	0.9219
146-175 กับ 176-208	0.1150	0.0000*	0.0006*	0.0000*
176-208 กับ 209-238	0.0000*	0.0066*	0.2759	0.0088*
209-238 กับ 239-297	0.0000*	0.0013*	0.0025*	0.0195*
239-297 กับ 298-332	0.0000*	0.0000*	0.0106*	0.1303
298-332 กับ 333-380	0.0000*	0.0006*	0.0000*	0.1027
333-380 กับ 381-425	0.0000*	0.0000*	0.0000*	0.0003*
381-425 กับ 426-455	0.0000*	0.0000*	0.0000*	0.0000*
426-455 กับ 456-499	0.0000*	0.0000*	0.0002*	0.0035*
456-499 กับ 500-551	0.0000*	0.0007*	0.0011*	0.0000*
500-551 กับ 552-581	0.0000*	0.0019*	0.0000*	0.0000*
552-581 กับ 582-627	0.0000*	0.0000*	0.0416*	0.7680
582-627 กับ 628-680	0.0000*	0.0000*	0.0001*	0.0010*
628-680 กับ 681-711	0.0000*	0.0000*	0.0000*	0.0000*

เปรียบเทียบช่วงวินาทีที่	ค่า P-Value			
	ค่าเฉลี่ยอัตราการเต้นของหัวใจ	ค่าเฉลี่ยชีพจร	ค่าความแปรปรวนของอัตราการเต้นของหัวใจ	ค่าความแปรปรวนของชีพจร
681-711 กับ 712-755	0.0000*	0.0000*	0.8815	0.0007*
712-755 กับ 756-785	0.0000*	0.0000*	0.0000*	0.2074
756-785 กับ 786-819	0.2789	0.1475	0.8045	0.0000*
786-819 กับ 820-854	0.1365	0.0000*	0.6038	0.5208
820-854 กับ 855-900	0.0000*	0.0000*	0.1180	0.9657

ตารางที่ 11 ค่า p-value การทดสอบสมมติฐานการเปลี่ยนแปลงค่าเฉลี่ย และค่าความแปรปรวนในแต่ละช่วง ที่ระดับนัยสำคัญ 0.05 เมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive หมายเหตุ ช่องที่มีเครื่องหมายดอกจัน (*) หมายถึงค่านั้นมีความต่างกันทางสถิติที่ระดับนัยสำคัญ 0.05

e-cp30



รูปที่ 20 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี e-cp30 โดยตรวจข้อมูลทั้งหมด

ผลจากการนำวิธี e-cp30 ไปใช้ในการตรวจหาจุดเปลี่ยนแปลงในข้อมูลสัญญาณชีพ พบว่าวิธี e-cp30 ตรวจพบจุดที่ข้อมูลมีการเปลี่ยนแปลงจำนวน 6 จุด ที่ตำแหน่ง 297, 432, 582, 682, 712 และ 750 ซึ่งมีค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลง ดังนี้

วินาทีที่	ค่าเฉลี่ยอัตราการเต้นของหัวใจ	ค่าเฉลี่ยชีพจร	ค่าความแปรปรวนของอัตราการเต้นของหัวใจ	ค่าความแปรปรวนของชีพจร	ค่าสหสัมพันธ์
1-296	88.64	87.85	1.91	23.47	0.23
297-431	92.34	92.33	6.44	6.21	0.80
432-581	80.21	80.63	5.24	4.84	0.84
582-681	69.33	69.05	3.78	6.33	0.94
682-711	72.27	72.20	6.48	7.13	0.75
712-749	65.82	65.29	1.18	1.56	0.52
750-900	76.40	75.96	1.44	4.37	0.69

ตารางที่ 12 ค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลงเมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี *e-cp3o*

และจากการนำข้อมูลในแต่ละช่วงไปทดสอบสมมติฐานความแตกต่างของค่าเฉลี่ย และค่าความแปรปรวนของสองประชากรโดยใช้สถิติทดสอบ *t* และ *F* ตามลำดับ พบว่า ในทุก ๆ ช่วงการเปรียบเทียบ ไม่ว่าจะเป็นการทดสอบความแตกต่างของค่าเฉลี่ยของอัตราการเต้นของหัวใจ ค่าเฉลี่ยของชีพจร ค่าความแปรปรวนของอัตราการเต้นของหัวใจ หรือค่าความแปรปรวนของชีพจร จะมีอย่างน้อยหนึ่งค่าที่มีค่า *p-value* น้อยกว่า 0.05 หมายถึงค่าระหว่างสองช่วงที่เปรียบเทียบกันนั้นมีความแตกต่างกันทางสถิติที่ระดับนัยสำคัญ 0.05 ซึ่งสอดคล้องกับตำแหน่งจุดเปลี่ยนแปลงที่พบด้วยวิธี *e-cp3o*

เปรียบเทียบช่วงวินาทีที่	ค่า P-Value			
	ค่าเฉลี่ยอัตราการเต้นของหัวใจ	ค่าเฉลี่ยชีพจร	ค่าความแปรปรวนของอัตราการเต้นของหัวใจ	ค่าความแปรปรวนของชีพจร
1-296 กับ 297-431	0.0000*	0.0000*	0.0000*	0.0000*
297-431 กับ 432-581	0.0000*	0.0000*	0.2228	0.1385
432-581 กับ 582-681	0.0000*	0.0000*	0.0807	0.1369
582-681 กับ 682-711	0.0000*	0.0000*	0.0530	0.6487

เปรียบเทียบช่วงวินาทีที่	ค่า P-Value			
	ค่าเฉลี่ย อัตราการ เต้นของ หัวใจ	ค่าเฉลี่ยชีพจร	ค่าความ แปรปรวนของ อัตราการเต้น ของหัวใจ	ค่าความ แปรปรวนของ ชีพจร
682-711 กับ 712-749	0.0000*	0.0000*	0.0000*	0.0000*
712-749 กับ 750-900	0.0000*	0.0000*	0.4848	0.0005*

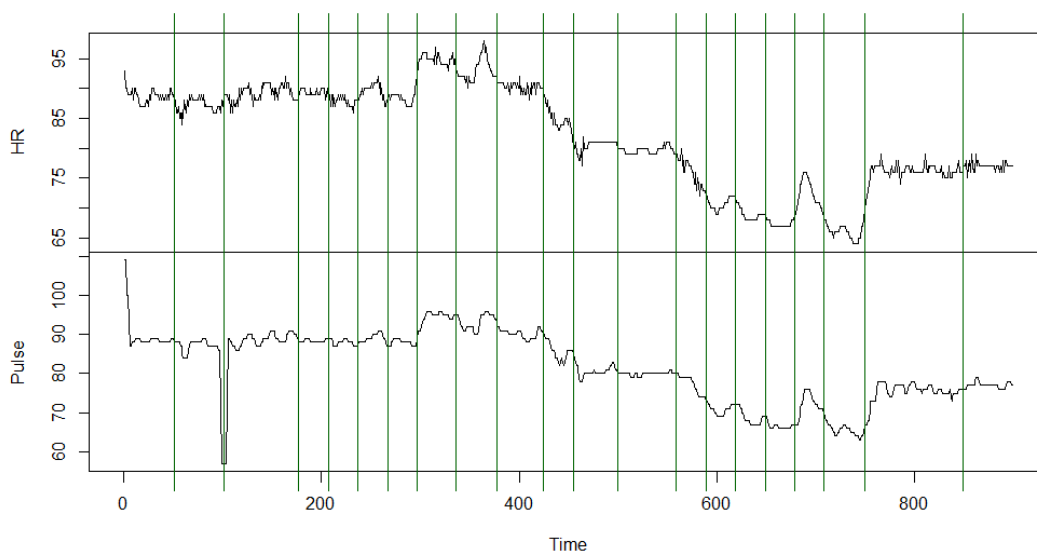
ตารางที่ 13 ค่า p-value การทดสอบสมมติฐานการเปลี่ยนแปลงค่าเฉลี่ย และค่าความแปรปรวนในแต่ละช่วง ที่ระดับนัยสำคัญ 0.05 เมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี e-cp30

หมายเหตุ ช่องที่มีเครื่องหมายดอกจัน (*) หมายถึงค่านั้นมีความต่างกันทางสถิติที่ระดับนัยสำคัญ

0.05



ks-cp30



รูปที่ 21 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี ks-cp30 โดยตรวจข้อมูลทั้งหมด

ผลจากการนำวิธี ks-cp30 ไปใช้ในการตรวจหาจุดเปลี่ยนแปลงในข้อมูลสัญญาณชีพ พบว่าวิธี ks-cp30 ตรวจพบจุดที่ข้อมูลมีการเปลี่ยนแปลงจำนวน 20 จุด ที่ตำแหน่ง 51, 101, 177, 207, 237, 267, 297, 336, 378, 425, 455, 500, 559, 589, 619, 649, 679, 709, 750 และ 850 ซึ่งมีค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลง ดังนี้

วินาทีที่	ค่าเฉลี่ยอัตรา การเต้นของ หัวใจ	ค่าเฉลี่ยชีพ จร	ค่าความ แปรปรวนของ อัตราการเต้น ของหัวใจ	ค่าความ แปรปรวนของ ชีพจร	ค่า สหสัมพันธ์
1-50	88.78	89.72	1.32	21.76	0.52
51-100	87.16	85.68	1.28	41.41	0.06
101-176	89.20	87.33	1.81	43.48	0.10
177-206	89.40	88.27	0.32	0.20	-0.30
207-236	87.90	87.97	0.64	0.65	0.15
237-266	89.63	89.13	1.55	1.43	0.54
267-296	88.40	87.90	0.94	0.58	-0.04
297-335	94.87	94.49	0.96	2.73	0.24
336-377	93.07	93.24	4.02	3.89	0.56
378-424	90.28	90.19	0.68	1.38	0.44
425-454	85.03	85.70	3.21	5.53	0.79
455-499	80.51	80.69	0.98	2.22	0.22
500-558	79.69	79.90	0.39	0.20	0.20
559-588	76.10	77.37	4.78	5.07	0.90
589-618	70.73	70.77	1.17	1.77	0.79
619-648	68.97	68.77	1.27	3.56	0.89
649-678	67.37	66.67	0.38	0.92	0.80
679-708	72.40	72.00	5.49	8.76	0.76
709-749	65.98	65.56	1.42	2.45	0.65
750-849	76.09	75.42	1.64	5.44	0.69
850-900	77.02	77.02	0.50	0.62	0.18

ตารางที่ 14 ค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลงเมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี ks-cp30

และจากการนำข้อมูลในแต่ละช่วงไปทดสอบสมมติฐานความแตกต่างของค่าเฉลี่ย และค่าความแปรปรวนของสองประชากรโดยใช้สถิติทดสอบ t และ F ตามลำดับ พบว่า ในทุก ๆ ช่วงการเปรียบเทียบ ไม่ว่าจะเป็นการทดสอบความแตกต่างของค่าเฉลี่ยของอัตราการเต้นของหัวใจ ค่าเฉลี่ยของชีพจร ค่าความแปรปรวนของอัตราการเต้นของหัวใจ หรือค่าความแปรปรวนของชีพจร จะมีอย่างน้อยหนึ่งค่าที่มีค่า p-value น้อยกว่า 0.05 หมายถึงค่าระหว่างสองช่วงที่เปรียบเทียบกันนั้นมีความแตกต่างกันทางสถิติที่ระดับนัยสำคัญ 0.05 ซึ่งสอดคล้องกับตำแหน่งจุดเปลี่ยนแปลงที่พบด้วยวิธี ks-cp3o

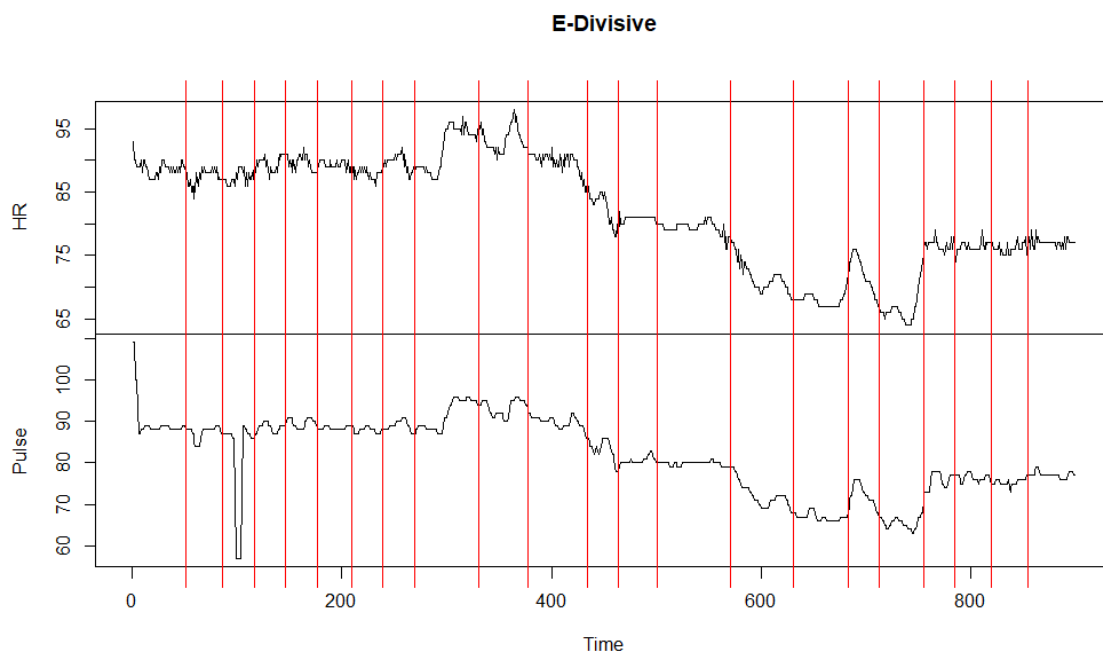
เปรียบเทียบช่วงวินาที ที่	ค่า P-Value			
	ค่าเฉลี่ยอัตราการ เต้นของ หัวใจ	ค่าเฉลี่ยชีพจร	ค่าความ แปรปรวนของ อัตราการเต้น ของหัวใจ	ค่าความ แปรปรวนของ ชีพจร
1-50 กับ 51-100	0.0000*	0.0005*	0.9189	0.0263*
51-100 กับ 101-176	0.0000*	0.1681	0.1939	0.8661
101-176 กับ 177-206	0.2774	0.2215	0.0000*	0.0000*
177-206 กับ 207-236	0.0000*	0.0825	0.0609	0.0023*
207-236 กับ 237-266	0.0000*	0.0001*	0.0211*	0.0392*
237-266 กับ 267-296	0.0001*	0.0000*	0.1819	0.0169*
267-296 กับ 297-335	0.0000*	0.0000*	0.9673	0.0000*
297-335 กับ 336-377	0.0000*	0.0029*	0.0000*	0.2722
336-377 กับ 378-424	0.0000*	0.0000*	0.0000*	0.0007*
378-424 กับ 425-454	0.0000*	0.0000*	0.0000*	0.0000*
425-454 กับ 455-499	0.0000*	0.0000*	0.0004*	0.0061*
455-499 กับ 500-558	0.0000*	0.0012*	0.0010*	0.0000*
500-558 กับ 559-588	0.0000*	0.0000*	0.0000*	0.0000*
559-588 กับ 589-618	0.0000*	0.0000*	0.0003*	0.0060*
589-618 กับ 619-648	0.0000*	0.0000*	0.8152	0.0645
619-648 กับ 649-678	0.0000*	0.0000*	0.0016*	0.0005*
649-678 กับ 679-708	0.0000*	0.0000*	0.0000*	0.0000*

เปรียบเทียบช่วงวินาที ที่	ค่า P-Value			
	ค่าเฉลี่ยอัตรา การเต้นของ หัวใจ	ค่าเฉลี่ยชีพจร	ค่าความ แปรปรวนของ อัตราการเต้น ของหัวใจ	ค่าความ แปรปรวนของ ชีพจร
679-708 กับ 709-749	0.0000*	0.0000*	0.0001*	0.0002*
709-749 กับ 750-849	0.0000*	0.0000*	0.6291	0.0056*
750-849 กับ 850-900	0.0000*	0.0000*	0.0000*	0.0000*

ตารางที่ 15 ค่า p-value การทดสอบสมมติฐานการเปลี่ยนแปลงค่าเฉลี่ย และค่าความแปรปรวนในแต่ละช่วง ที่ระดับนัยสำคัญ 0.05 เมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี ks-cp30
หมายเหตุ ช่องที่มีเครื่องหมายดอกจัน (*) หมายถึงค่านั้นมีความต่างกันทางสถิติที่ระดับนัยสำคัญ 0.05

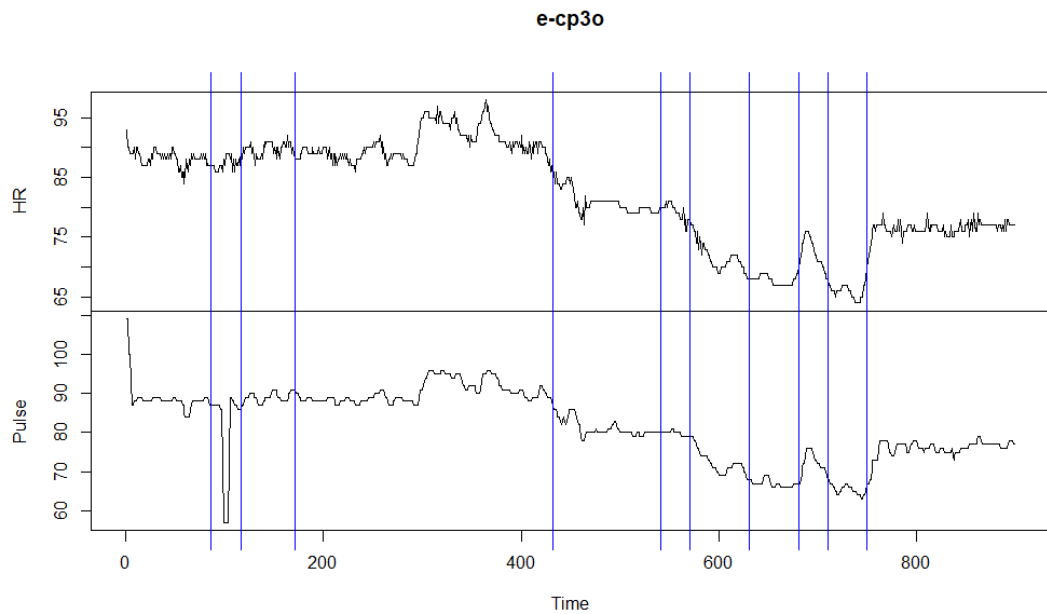
การเปรียบเทียบจำนวนจุดและตำแหน่งข้อมูลที่มีการเปลี่ยนแปลงที่ตรวจพบด้วยวิธี E-Divisive, e-cp30 และ ks-cp30 ในข้อมูลสัญญาณชีพ พบว่า วิธี E-Divisive และ ks-cp30 ตรวจพบจำนวนจุดที่มีการเปลี่ยนแปลงเป็นจำนวนใกล้เคียงกัน และวิธี e-cp30 ตรวจพบจำนวนจุดที่มีการเปลี่ยนแปลงน้อยที่สุด

4.2.2 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงของข้อมูลสัญญาณชีพที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o โดยทำการแบ่งข้อมูลออกเป็น 3 ส่วนก่อน

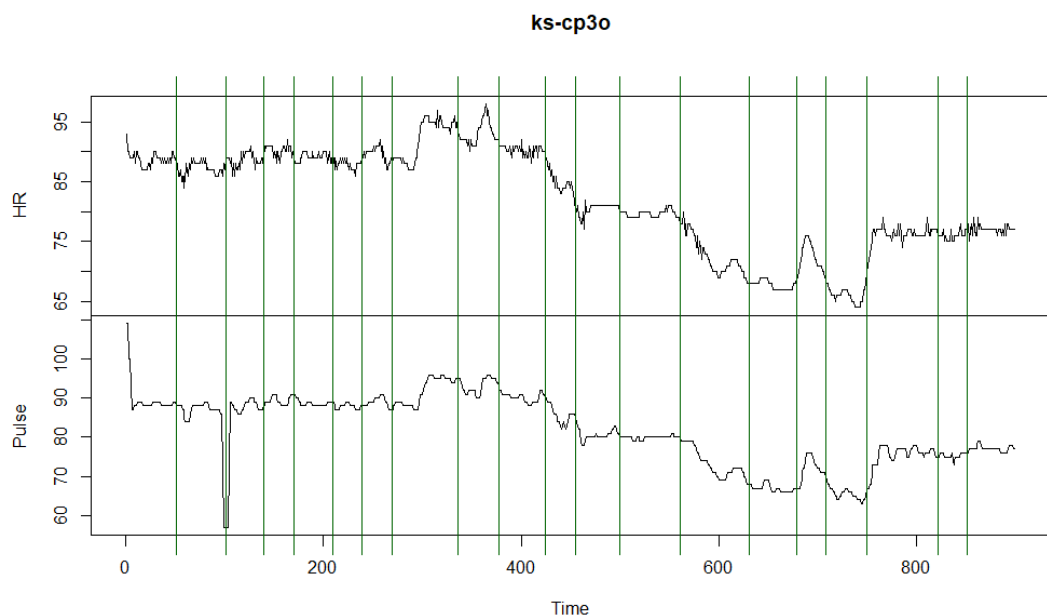


รูปที่ 22 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive โดยทำการแบ่งข้อมูลก่อน

ผลจากการนำวิธี E-Divisive ไปใช้ในการตรวจหาจุดเปลี่ยนแปลงในข้อมูลสัญญาณชีพ พบว่าวิธี E-Divisive ตรวจพบจุดที่ข้อมูลมีการเปลี่ยนแปลงจำนวน 21 จุด ที่ตำแหน่ง 51, 86, 116, 146, 176, 209, 239, 269, 331, 378, 434, 464, 501, 541, 571, 631, 683, 713, 755, 785, 820 และ 855



รูปที่ 23 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี e-cp30 โดยทำการแบ่งข้อมูลก่อน
 ผลจากการนำวิธี e-cp30 ไปใช้ในการตรวจหาจุดเปลี่ยนแปลงในข้อมูลสัญญาณชีพ พบว่า
 วิธี e-cp30 ตรวจพบจุดที่ข้อมูลมีการเปลี่ยนแปลงจำนวน 10 จุด ที่ตำแหน่ง 86, 116, 171, 432,
 541, 571, 631, 681, 711 และ 750



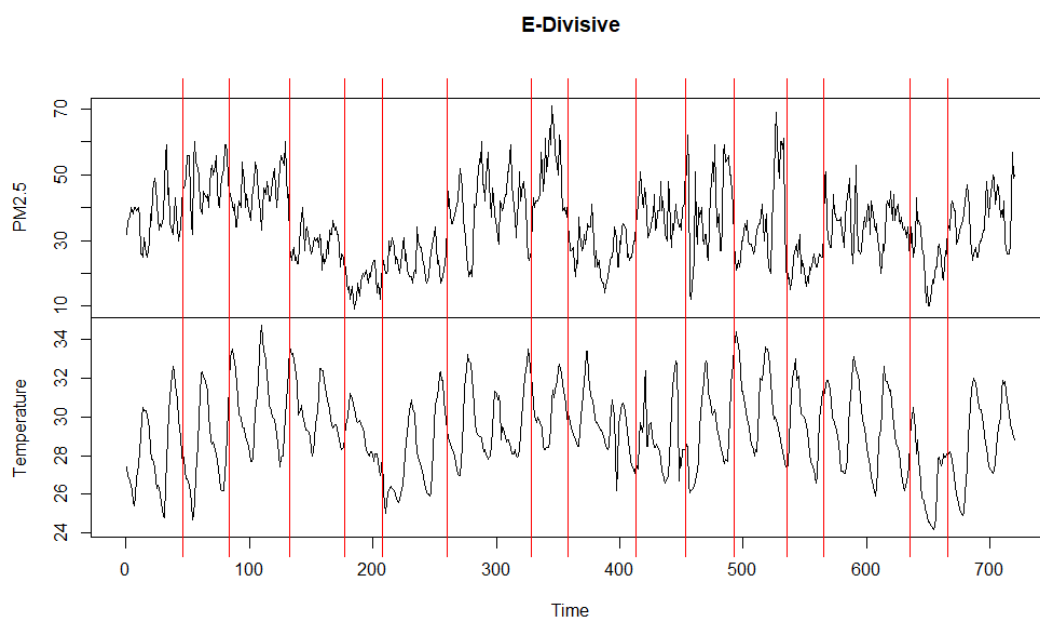
รูปที่ 24 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี ks-cp30 โดยทำการแบ่งข้อมูลก่อน

ผลจากการนำวิธี ks-cp3o ไปใช้ในการตรวจหาจุดเปลี่ยนแปลงในข้อมูลสัญญาณชีพ พบว่าวิธี ks-cp3o ตรวจพบจุดที่ข้อมูลมีการเปลี่ยนแปลงจำนวน 19 จุด ที่ตำแหน่ง 51, 101, 139, 170, 209, 239, 269, 336, 378, 425, 455, 500, 561, 631, 679, 709, 750, 822 และ 852

การเปรียบเทียบจำนวนจุดและตำแหน่งข้อมูลที่มีการเปลี่ยนแปลงที่ตรวจพบด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o ในข้อมูลสัญญาณชีพโดยแบ่งข้อมูลออกเป็น 3 ส่วนก่อนนำไปตรวจหาจุดเปลี่ยนแปลง พบว่า ให้ผลเช่นเดียวกับการตรวจหาจุดเปลี่ยนแปลงในข้อมูลทั้งหมดในคราวเดียว คือ วิธี E-Divisive และ ks-cp3o ตรวจพบจำนวนจุดที่มีการเปลี่ยนแปลงเป็นจำนวนใกล้เคียงกัน และวิธี e-cp3o ตรวจพบจำนวนจุดที่มีการเปลี่ยนแปลงน้อยที่สุด โดยการตรวจหาด้วยวิธี e-cp3o พบจำนวนจุดที่มีการเปลี่ยนแปลงมากขึ้น และ วิธี E-Divisive และ ks-cp3o พบจำนวนจุดที่มีการเปลี่ยนแปลงน้อยลง

การเปรียบเทียบการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อทำการตรวจหาข้อมูลทั้งหมดในคราวเดียวและแบ่งข้อมูลออกเป็น 3 ส่วนก่อน พบว่า การตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive และ ks-cp3o มีผลการตรวจหาที่ใกล้เคียงกัน แต่วิธี e-cp3o มีผลที่แตกต่างออกไปคือ ถ้าตรวจหาข้อมูลทั้งหมดในคราวเดียว จะไม่พบจุดเปลี่ยนแปลงในช่วงวินาทีที่ 0 ถึง 200 แต่ถ้าแบ่งข้อมูลออกเป็น 3 ส่วนก่อนที่จะนำไปตรวจหาจุดเปลี่ยนแปลงจะพบจุดที่มีการเปลี่ยนแปลงในช่วงวินาทีดังกล่าว ในทางกลับกัน ในช่วงวินาทีที่ 200 ถึง 400 ถ้าตรวจหาจุดเปลี่ยนแปลงทั้งหมดในคราวเดียวจะพบจุดเปลี่ยนแปลงในขณะที่ถ้าแบ่งข้อมูลออกเป็น 3 ส่วนก่อนจะไม่พบ

4.2.3 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงของข้อมูลปริมาณฝุ่นละออง PM2.5 ที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o โดยทำการตรวจหาข้อมูลทั้งหมดในคราวเดียว



รูปที่ 25 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive โดยตรวจข้อมูลทั้งหมด

ผลจากการนำวิธี E-Divisive ไปใช้ในการตรวจหาจุดเปลี่ยนแปลงในข้อมูลสัญญาณปริมาณฝุ่นละออง PM2.5 พบว่า วิธี E-Divisive ตรวจพบจุดที่ข้อมูลมีการเปลี่ยนแปลงจำนวน 15 จุด ที่ตำแหน่ง 46, 84, 133, 177, 208, 260, 328, 358, 413, 453, 493, 535, 565, 635 และ 666 ซึ่งมีค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลง ดังนี้

ชั่วโมงที่	ค่าเฉลี่ยความหนาแน่นฝุ่นละออง PM2.5	ค่าเฉลี่ยอุณหภูมิ	ค่าความแปรปรวนของความหนาแน่นฝุ่นละออง PM2.5	ค่าความแปรปรวนของอุณหภูมิ	ค่าสหสัมพันธ์
1-45	37.2222	28.2889	56.6768	4.6796	-0.3789
46-83	47.9211	28.2868	48.3990	4.6633	-0.2500
84-132	44.6327	30.5673	36.8622	4.0218	-0.3884
133-176	29.0682	30.3727	18.8557	2.5104	-0.2613
177-207	17.2258	29.0323	16.1806	1.3949	-0.4661

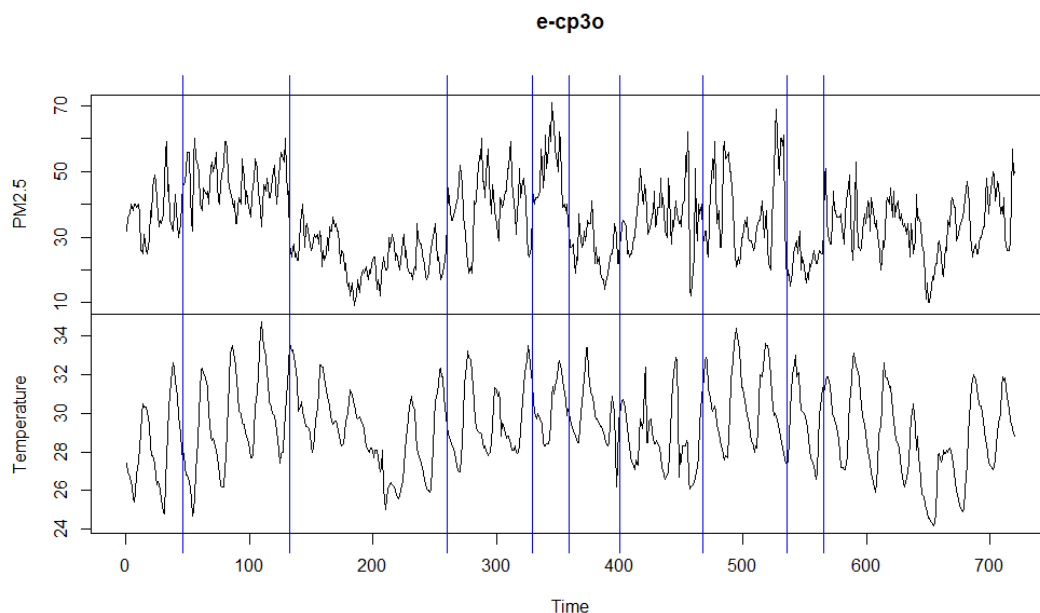
ชั่วโมงที่	ค่าเฉลี่ยความ หนาแน่นฝุ่น ละออง PM2.5	ค่าเฉลี่ย อุณหภูมิ	ค่าความ แปรปรวนของ ความหนาแน่น ฝุ่นละออง PM2.5	ค่าความ แปรปรวนของ อุณหภูมิ	ค่า สหสัมพันธ์
208-259	23.9038	28.0231	21.5004	4.6300	-0.2199
260-327	40.4853	29.5647	89.3580	3.0823	-0.6439
328-357	50.0333	30.4133	112.4471	1.9171	-0.0745
358-412	27.0545	29.6600	36.6822	2.1113	0.4367
413-452	38.1750	28.8675	29.0712	3.0120	0.0469
453-492	40.4500	29.2525	174.3051	4.2000	0.0556
493-534	36.2143	30.7024	164.3676	4.1661	-0.4689
535-564	22.3000	29.5900	18.8379	3.5313	0.3356
565-634	35.2286	29.3929	44.4398	4.2186	-0.0232
635-665	24.1613	26.9032	69.4065	3.5970	0.6611
666-720	37.2000	28.7673	64.0148	4.3019	-0.3661

ตารางที่ 16 ค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลงเมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive

และจากการนำข้อมูลในแต่ละช่วงไปทดสอบสมมติฐานความแตกต่างของค่าเฉลี่ย และค่าความแปรปรวนของสองประชากรโดยใช้สถิติทดสอบ t และ F ตามลำดับ พบว่า โดยส่วนใหญ่ในทุก ๆ ช่วงการเปรียบเทียบ ไม่ว่าจะเป็นการทดสอบความแตกต่างของค่าเฉลี่ยของอัตราการเต้นของหัวใจ ค่าเฉลี่ยของชีพจร ค่าความแปรปรวนของอัตราการเต้นของหัวใจ หรือค่าความแปรปรวนของชีพจร จะมีอย่างน้อยหนึ่งค่าที่มีค่า p-value น้อยกว่า 0.05 หมายถึงค่าระหว่างสองช่วงที่เปรียบเทียบกันนั้น มีความแตกต่างกัน ซึ่งสอดคล้องกับตำแหน่งจุดเปลี่ยนแปลงที่พบด้วยวิธี E-Divisive

เปรียบเทียบช่วงชั่วโมง ที่	ค่า P-Value			
	ค่าเฉลี่ย ความ หนาแน่น ฝุ่นละออง PM2.5	ค่าเฉลี่ย อุณหภูมิ	ค่าความ แปรปรวนของ ความหนาแน่น ฝุ่นละออง PM2.5	ค่าความ แปรปรวนของ อุณหภูมิ
1-45 กับ 46-83	0.0000*	0.9966	0.6263	0.9985
46-83 กับ 84-132	0.0211*	0.0000*	0.3723	0.6240
84-132 กับ 133-176	0.0000*	0.6076	0.0274*	0.1190
133-176 กับ 177-207	0.0000*	0.0002*	0.6677	0.0942
177-207 กับ 208-259	0.0000*	0.0072*	0.4077	0.0007*
208-259 กับ 260-327	0.0000*	0.0000*	0.0000*	0.1182
260-327 กับ 328-357	0.0000*	0.0212*	0.4359	0.1586
328-357 กับ 358-412	0.0000*	0.0227*	0.0004*	0.7946
358-412 กับ 413-452	0.0000*	0.0176*	0.4493	0.2250
413-452 กับ 453-492	0.3177	0.3674	0.0000*	0.3034
453-492 กับ 493-534	0.1444	0.0019*	0.8516	0.9776
493-534 กับ 535-564	0.0000*	0.0213*	0.0000*	0.6483
535-564 กับ 565-634	0.0000*	0.6531	0.0122*	0.6063
565-634 กับ 635-665	0.0000*	0.0000*	0.1305	0.6409
635-665 กับ 666-720	0.0000*	0.0001*	0.7785	0.6055

ตารางที่ 17 ค่า p-value การทดสอบสมมติฐานการเปลี่ยนแปลงค่าเฉลี่ย และค่าความแปรปรวนในแต่ละช่วง ที่ระดับนัยสำคัญ 0.05 เมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive
หมายเหตุ ช่องที่มีเครื่องหมายดอกจัน (*) หมายถึงค่านั้นมีความต่างกันทางสถิติที่ระดับนัยสำคัญ 0.05



รูปที่ 26 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี e-cp30 โดยตรวจข้อมูลทั้งหมด

ผลจากการนำวิธี e-cp30 ไปใช้ในการตรวจหาจุดเปลี่ยนแปลงในข้อมูลสัญญาณปริมาณฝุ่นละออง PM2.5 พบว่า วิธี e-cp30 ตรวจพบจุดที่ข้อมูลมีการเปลี่ยนแปลงจำนวน 9 จุด ที่ตำแหน่ง 46, 133, 260, 329, 359, 400, 467, 535 และ 565 ซึ่งมีค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลง ดังนี้

ชั่วโมงที่	ค่าเฉลี่ยความหนาแน่นฝุ่นละออง PM2.5	ค่าเฉลี่ยอุณหภูมิ	ค่าความแปรปรวนของความหนาแน่นฝุ่นละออง PM2.5	ค่าความแปรปรวนของอุณหภูมิ	ค่าสหสัมพันธ์
1-45	37.2222	28.2889	56.6768	4.6796	-0.3789
46-132	46.0690	29.5713	44.0882	5.5453	-0.3915
133-259	24.0630	29.0835	39.2500	4.1081	0.0300
260-328	40.2899	29.6000	90.6795	3.1229	-0.6539
329-358	50.2667	30.3533	102.9609	1.8281	0.0218
359-399	26.1951	29.8415	41.3610	2.1560	0.5000
400-466	36.1940	28.6299	80.8557	2.8097	0.1692

ชั่วโมงที่	ค่าเฉลี่ยความ หนาแน่นฝุ่น ละออง PM2.5	ค่าเฉลี่ย อุณหภูมิ	ค่าความ แปรปรวนของ ความหนาแน่น ฝุ่นละออง PM2.5	ค่าความ แปรปรวนของ อุณหภูมิ	ค่า สหสัมพันธ์
467-534	38.5441	30.4956	159.0577	3.7551	-0.4680
535-564	22.3000	29.5900	18.8379	3.5313	0.3356
565-720	33.7244	28.6776	79.1171	4.9365	0.1900

ตารางที่ 18 ค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลงเมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี e-cp3o

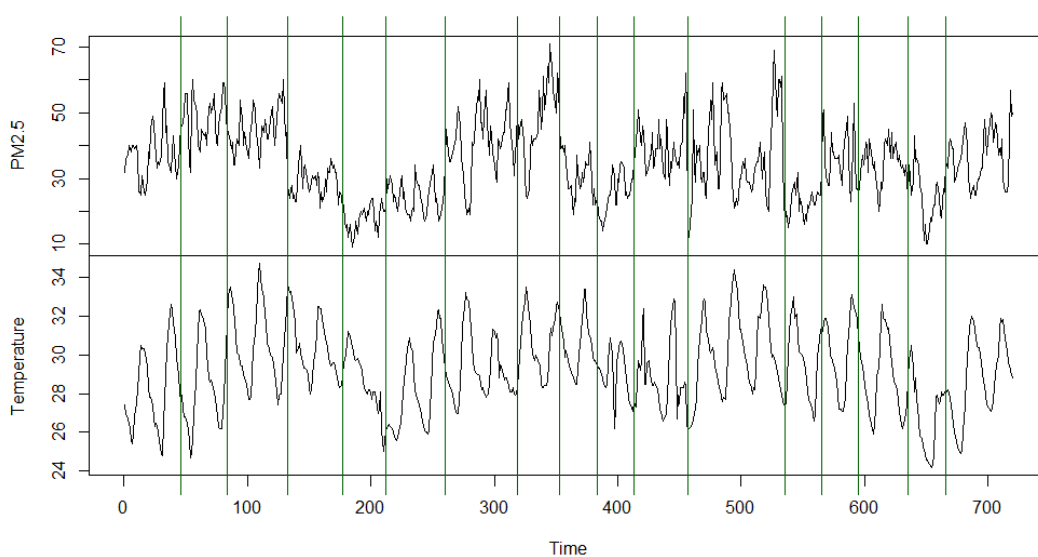
และจากการนำข้อมูลในแต่ละช่วงไปทดสอบสมมติฐานความแตกต่างของค่าเฉลี่ย และค่าความแปรปรวนของสองประชากรโดยใช้สถิติทดสอบ t และ F ตามลำดับ พบว่า โดยส่วนใหญ่ในทุก ๆ ช่วงการเปรียบเทียบ ไม่ว่าจะเป็นการทดสอบความแตกต่างของค่าเฉลี่ยของอัตราการเต้นของหัวใจ ค่าเฉลี่ยของชีพจร ค่าความแปรปรวนของอัตราการเต้นของหัวใจ หรือค่าความแปรปรวนของชีพจร จะมีอย่างน้อยหนึ่งค่าที่มีค่า p-value น้อยกว่า 0.05 หมายถึงค่าระหว่างสองช่วงที่เปรียบเทียบกันนั้น มีความแตกต่างกัน ซึ่งสอดคล้องกับตำแหน่งจุดเปลี่ยนแปลงที่พบด้วยวิธี e-cp3o

เปรียบเทียบช่วงชั่วโมง ที่	ค่า P-Value			
	ค่าเฉลี่ย ความ หนาแน่น ฝุ่นละออง PM2.5	ค่าเฉลี่ย อุณหภูมิ	ค่าความ แปรปรวนของ ความหนาแน่น ฝุ่นละออง PM2.5	ค่าความ แปรปรวนของ อุณหภูมิ
1-45 กับ 46-132	0.0000*	0.0028*	0.3200	0.5416
46-132 กับ 133-259	0.0000*	0.1071	0.5481	0.1243
133-259 กับ 260-328	0.0000*	0.0766	0.0000*	0.2133
260-328 กับ 329-358	0.0000*	0.0399*	0.6540	0.1119
329-358 กับ 359-399	0.0000*	0.1383	0.0078*	0.6498
359-399 กับ 400-466	0.0000*	0.0002*	0.0245*	0.3704
400-466 กับ 467-534	0.2144	0.0000*	0.0065*	0.2400

เปรียบเทียบช่วงชั่วโมง ที่	ค่า P-Value			
	ค่าเฉลี่ย ความ หนาแน่น ฝุ่นละออง PM2.5	ค่าเฉลี่ย อุณหภูมิ	ค่าความ แปรปรวนของ ความหนาแน่น ฝุ่นละออง PM2.5	ค่าความ แปรปรวนของ อุณหภูมิ
467-534 กับ 535-564	0.0000*	0.0339*	0.0000*	0.8792
535-564 กับ 565-720	0.0000*	0.0364*	0.0000*	0.2906

ตารางที่ 19 ค่า p-value การทดสอบสมมติฐานการเปลี่ยนแปลงค่าเฉลี่ย และค่าความแปรปรวนในแต่ละช่วง ที่ระดับนัยสำคัญ 0.05 เมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี $e-cp3o$ หมายถึง ช่องที่มีเครื่องหมายดอกจัน (*) หมายถึงค่านั้นมีความต่างกันทางสถิติที่ระดับนัยสำคัญ 0.05

ks-cp3o



รูปที่ 27 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี $ks-cp3o$ โดยตรวจข้อมูลทั้งหมด

ผลจากการนำวิธี $ks-cp3o$ ไปใช้ในการตรวจหาจุดเปลี่ยนแปลงในข้อมูลสัญญาณปริมาณฝุ่นละออง PM2.5 พบว่า วิธี $ks-cp3o$ ตรวจพบจุดที่ข้อมูลมีการเปลี่ยนแปลงจำนวน 16 จุด ที่ตำแหน่ง 46, 84, 133, 177, 212, 260, 319, 353, 383, 413, 457, 535, 565, 595, 635 และ 666 ซึ่งมีค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลง ดังนี้

ชั่วโมงที่	ค่าเฉลี่ยความ หนาแน่นฝุ่น ละออง PM2.5	ค่าเฉลี่ย อุณหภูมิ	ค่าความ แปรปรวนของ ความหนาแน่น ฝุ่นละออง PM2.5	ค่าความ แปรปรวนของ อุณหภูมิ	ค่า สหสัมพันธ์
1-45	37.2222	28.2889	56.6768	4.6796	-0.3789
46-83	47.9211	28.2868	48.3990	4.6633	-0.2500
84-132	44.6327	30.5673	36.8622	4.0218	-0.3884
133-176	29.0682	30.3727	18.8557	2.5104	-0.2613
177-211	17.6857	28.7086	16.2807	2.2261	-0.5145
212-259	24.1250	28.1750	22.4521	4.6023	-0.2812
260-318	40.6610	29.2746	88.9176	2.4668	-0.6781
319-352	48.5588	30.5912	138.0722	2.5723	-0.3395
353-382	31.2000	30.4767	45.7517	1.8825	0.4943
383-412	25.2000	29.0833	35.5448	1.4821	0.2266
413-456	39.6591	28.7864	50.9741	2.8677	-0.0566
457-534	37.4487	30.0987	165.0558	4.6438	-0.2059
535-564	22.3000	29.5900	18.8379	3.5313	0.3356
565-594	36.0333	30.2000	61.5506	3.7062	-0.0740
595-634	34.6250	28.7875	31.9840	3.8309	-0.0546
635-665	24.1613	26.9032	69.4065	3.5970	0.6611
666-720	37.2000	28.7673	64.0148	4.3019	-0.3661

ตารางที่ 20 ค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ในแต่ละช่วงการเปลี่ยนแปลงเมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี ks-cp3o

และจากการนำข้อมูลในแต่ละช่วงไปทดสอบสมมติฐานความแตกต่างของค่าเฉลี่ย และค่าความแปรปรวนของสองประชากรโดยใช้สถิติทดสอบ t และ F ตามลำดับ พบว่า ในทุก ๆ ช่วงการเปรียบเทียบ ไม่ว่าจะเป็นการทดสอบความแตกต่างของค่าเฉลี่ยของอัตราการเต้นของหัวใจ ค่าเฉลี่ยของชีพจร ค่าความแปรปรวนของอัตราการเต้นของหัวใจ หรือค่าความแปรปรวนของชีพจร จะมีอย่างน้อยหนึ่งค่าที่มีค่า p-value น้อยกว่า 0.05 หมายถึงค่าระหว่างสองช่วงที่เปรียบเทียบกันนั้นมีความแตกต่างกัน ซึ่งสอดคล้องกับตำแหน่งจุดเปลี่ยนแปลงที่พบด้วยวิธี ks-cp3o

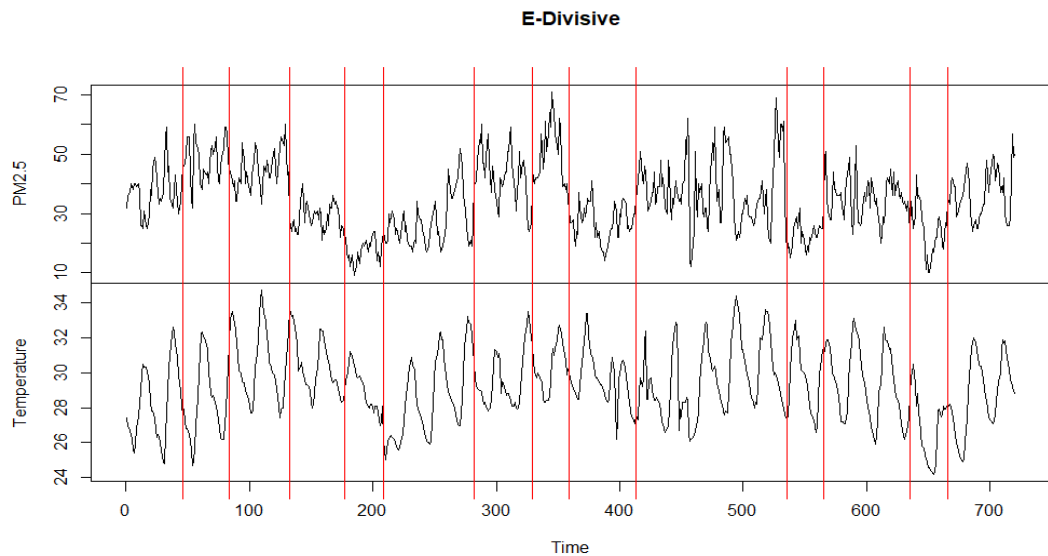
เปรียบเทียบช่วงชั่วโมง ที่	ค่า P-Value			
	ค่าเฉลี่ย ความ หนาแน่น ฝุ่นละออง PM2.5	ค่าเฉลี่ย อุณหภูมิ	ค่าความ แปรปรวนของ ความหนาแน่น ฝุ่นละออง PM2.5	ค่าความ แปรปรวนของ อุณหภูมิ
1-45 กับ 46-83	0.0000*	0.9966	0.6263	0.9985
46-83 กับ 84-132	0.0211*	0.0000*	0.3723	0.6240
84-132 กับ 133-176	0.0000*	0.6076	0.0274*	0.1190
133-176 กับ 177-211	0.0000*	0.0000*	0.6633	0.7231
177-211 กับ 212-259	0.0000*	0.1853	0.3293	0.0289*
212-259 กับ 260-318	0.0000*	0.0040*	0.0000*	0.0243*
260-318 กับ 319-352	0.0006*	0.0002*	0.1413	0.8699
319-352 กับ 353-382	0.0000*	0.7616	0.0034*	0.3964
353-382 กับ 383-412	0.0006*	0.0001*	0.5010	0.5237
383-412 กับ 413-456	0.0000*	0.4119	0.3093	0.0641
413-456 กับ 457-534	0.2243	0.0007*	0.0001*	0.0872
457-534 กับ 535-564	0.0000*	0.2582	0.0000*	0.4130
535-564 กับ 565-594	0.0000*	0.2193	0.0021*	0.8973
565-594 กับ 595-634	0.3856	0.0037*	0.0568	0.9382
595-634 กับ 635-665	0.0000*	0.0001*	0.0235*	0.8675
635-665 กับ 666-720	0.0000*	0.0001*	0.7785	0.6055

ตารางที่ 21 ค่า p-value การทดสอบสมมติฐานการเปลี่ยนแปลงค่าเฉลี่ย และค่าความแปรปรวนในแต่ละช่วง ที่ระดับนัยสำคัญ 0.05 เมื่อตรวจหาจุดเปลี่ยนแปลงด้วยวิธี ks-cp30

หมายเหตุ ช่องที่มีเครื่องหมายดอกจัน (*) หมายถึงค่านั้นมีความต่างกันทางสถิติที่ระดับนัยสำคัญ

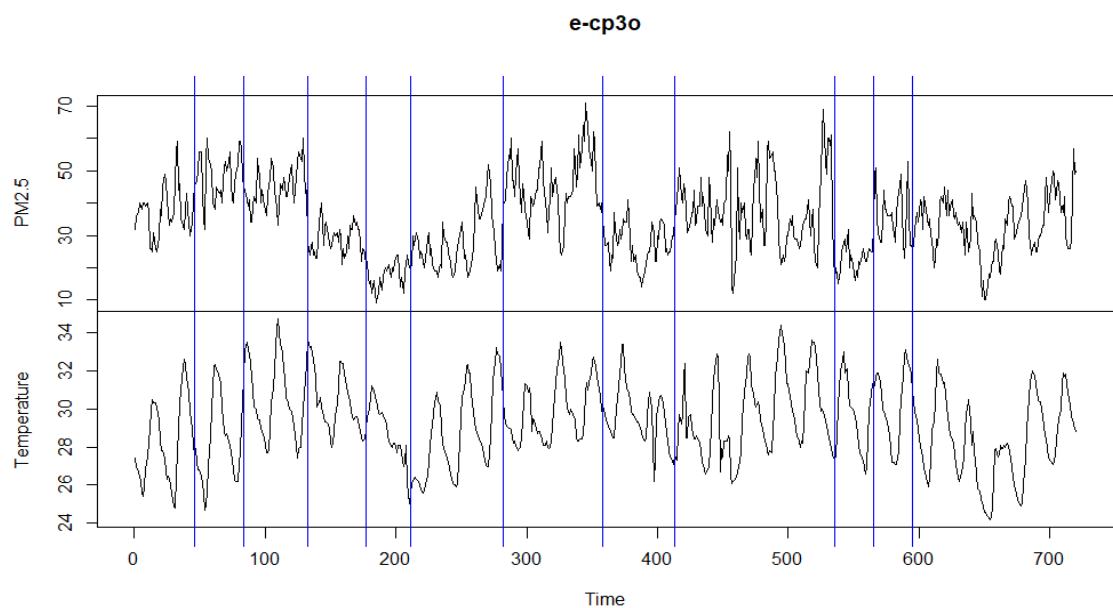
0.05

4.2.4 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงของข้อมูลปริมาณฝุ่นละออง PM2.5 ที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o โดยทำการแบ่งข้อมูลออกเป็น 3 ส่วนก่อน



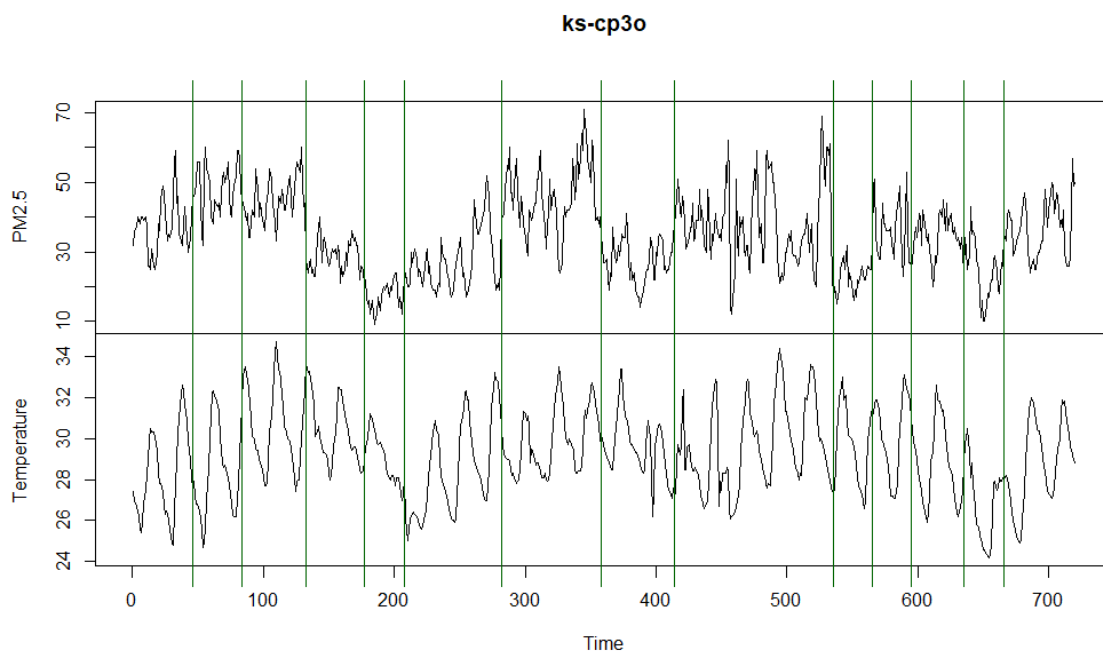
รูปที่ 28 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive โดยทำการแบ่งข้อมูลก่อน

ผลจากการนำวิธี E-Divisive ไปใช้ในการตรวจหาจุดเปลี่ยนแปลงในข้อมูลปริมาณฝุ่นละออง PM2.5 พบว่า วิธี E-Divisive ตรวจพบจุดที่ข้อมูลมีการเปลี่ยนแปลงจำนวน 13 จุด ที่ตำแหน่ง 46, 84, 133, 177, 209, 282, 329, 359, 413, 535, 565, 635 และ 666



รูปที่ 29 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี e-cp3o โดยทำการแบ่งข้อมูลก่อน

ผลจากการนำวิธี e-cp3o ไปใช้ในการตรวจหาจุดเปลี่ยนแปลงในข้อมูลปริมาณฝุ่นละออง PM2.5 พบว่า วิธี e-cp3o ตรวจพบจุดที่ข้อมูลมีการเปลี่ยนแปลงจำนวน 11 จุด ที่ตำแหน่ง 46, 84, 133, 177, 211, 282, 358, 413, 535, 565 และ 595



รูปที่ 30 ตำแหน่งจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี ks-cp3o โดยทำการแบ่งข้อมูลก่อน

ผลจากการนำวิธี ks-cp3o ไปใช้ในการตรวจหาจุดเปลี่ยนแปลงในข้อมูลปริมาณฝุ่นละออง PM2.5 พบว่า วิธี ks-cp3o ตรวจพบจุดที่ข้อมูลมีการเปลี่ยนแปลงจำนวน 13 จุด ที่ตำแหน่ง 46, 84, 133, 177, 208, 282, 358, 414, 535, 565, 595, 635 และ 666

การเปรียบเทียบจำนวนจุดและตำแหน่งข้อมูลที่มีการเปลี่ยนแปลงที่ตรวจพบด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o ในข้อมูลสัญญาณชีพโดยแบ่งข้อมูลออกเป็น 3 ส่วนก่อนนำไปตรวจหาจุดเปลี่ยนแปลง พบว่า ให้ผลเช่นเดียวกับการตรวจหาจุดเปลี่ยนแปลงในข้อมูลทั้งหมดในคราวเดียว คือ วิธี E-Divisive และ ks-cp3o ตรวจพบจำนวนจุดที่มีการเปลี่ยนแปลงเป็นจำนวนใกล้เคียงกัน และวิธี e-cp3o ตรวจพบจำนวนจุดที่มีการเปลี่ยนแปลงน้อยที่สุด โดยการตรวจหาด้วยวิธี e-cp3o พบจำนวนจุดที่มีการเปลี่ยนแปลงมากขึ้น และ วิธี E-Divisive และ ks-cp3o พบจำนวนจุดที่มีการเปลี่ยนแปลงน้อยลง

การเปรียบเทียบการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อทำการตรวจหาข้อมูลทั้งหมดในคราวเดียวและแบ่งข้อมูลออกเป็น 3 ส่วนก่อน พบว่า ทุกวิธีให้

จำนวนและตำแหน่งจุดเปลี่ยนแปลงใกล้เคียงกันทั้งหมดไม่ว่าจะตรวจหาข้อมูลทั้งหมดในคราวเดียว
หรือแบ่งข้อมูลออกเป็น 3 ส่วนก่อน



บทที่ 5

สรุปผลวิจัยและข้อเสนอแนะ

การศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบของวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลอนุกรมเวลาที่มีความผันแปรไม่ปกติที่มีการแจกแจงปกติ 2 ตัวแปรมีการเปลี่ยนแปลงในค่าเฉลี่ย ค่าความแปรปรวน และค่าสหสัมพันธ์ โดยทำการศึกษาเปรียบเทียบใน ส่วนข้อมูลจำลอง จากนั้นทำการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงโดย พิจารณาจากค่า Adjusted Rand Index ซึ่งวิธีที่มีประสิทธิภาพมากที่สุดคือวิธีที่มีค่า Adjusted Rand Index สูงที่สุด และเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบโดยพิจารณาจากจำนวนจุด เปลี่ยนแปลงเฉลี่ยที่พบ ซึ่งวิธีที่แม่นยำที่สุดคือวิธีที่มีค่าเฉลี่ยจุดเปลี่ยนแปลงที่พบใกล้เคียง 2 มาก ที่สุดเนื่องจากข้อมูลที่จำลองขึ้นมานั้นมีจุดเปลี่ยนแปลงที่แท้จริงคือ 2 จุด และในส่วนข้อมูลจริง ซึ่งเป็นข้อมูลสัญญาณชีพของมหาวิทยาลัยควีนส์แลนด์ (The university of Queensland) โดยมีตัว แปรที่สนใจคือ อัตราการเต้นของหัวใจ (Heart rate) และอัตราการเต้นของชีพจร (Pulse) ซึ่งมีขนาด 900 วินาที (นาที่ที่ 0.12 ถึง นาที่ที่ 15.11) และข้อมูลปริมาณฝุ่นละออง PM2.5 จาก BERKELEY EARTH ซึ่งมีขนาด 720 จุด (ข้อมูลของวันที่ 1 พฤศจิกายน 2561 ถึง 30 พฤศจิกายน 2561) ซึ่งมีตัว แปรที่สนใจคือ ความหนาแน่นของปริมาณฝุ่นละออง PM2.5 และอุณหภูมิ โดยทำการตรวจหาข้อมูล ทั้งหมดในคราวเดียว และโดยการแบ่งข้อมูลออกเป็น 3 ส่วนก่อนแล้วจึงนำแต่ละส่วนมาตรวจหาจุด เปลี่ยนแปลง ซึ่งสรุปผลการวิจัยได้ดังนี้

5.1 สรุปผลการวิจัย

5.1.1 ผลการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 90

ข้อมูลที่มีการเปลี่ยนแปลงใน	อันดับประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลง
ค่าเฉลี่ย	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive
ค่าความแปรปรวน	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive
ค่าสหสัมพันธ์ ($\rho = 0.5$)	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive

ข้อมูลมีการเปลี่ยนแปลงใน	อันดับประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลง
ค่าสหสัมพันธ์ ($\rho = 0.9$)	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive
ค่าเฉลี่ย และค่าความแปรปรวน	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive
ค่าความแปรปรวน และค่าสหสัมพันธ์	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive
ค่าเฉลี่ย และค่าสหสัมพันธ์	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive
ค่าเฉลี่ย, ค่าความแปรปรวน และค่าสหสัมพันธ์	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive

ตารางที่ 22 ตารางแสดงผลการเปรียบเทียบประสิทธิภาพวิธีการตรวจหาจุดเปลี่ยนแปลงเมื่อข้อมูลมีขนาด 90

5.1.2 ผลการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 150

ข้อมูลมีการเปลี่ยนแปลงใน	อันดับประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลง
ค่าเฉลี่ย	อันดับที่ 1 วิธี E-Divisive อันดับที่ 2 วิธี e-cp3o อันดับที่ 3 วิธี ks-cp3o
ค่าความแปรปรวน	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี ks-cp3o อันดับที่ 3 วิธี E-Divisive
ค่าสหสัมพันธ์ ($\rho = 0.5$)	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี ks-cp3o อันดับที่ 3 วิธี E-Divisive
ค่าสหสัมพันธ์ ($\rho = 0.9$)	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี ks-cp3o อันดับที่ 3 วิธี E-Divisive

ข้อมูลมีการเปลี่ยนแปลงใน	อันดับประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลง
ค่าเฉลี่ย และค่าความแปรปรวน	อันดับที่ 1 วิธี E-Divisive อันดับที่ 2 วิธี e-cp3o อันดับที่ 3 วิธี ks-cp3o
ค่าความแปรปรวน และค่าสหสัมพันธ์	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี ks-cp3o อันดับที่ 3 วิธี E-Divisive
ค่าเฉลี่ย และค่าสหสัมพันธ์	อันดับที่ 1 วิธี E-Divisive อันดับที่ 2 วิธี e-cp3o อันดับที่ 3 วิธี ks-cp3o
ค่าเฉลี่ย, ค่าความแปรปรวน และค่าสหสัมพันธ์	อันดับที่ 1 วิธี E-Divisive อันดับที่ 2 วิธี e-cp3o อันดับที่ 3 วิธี ks-cp3o

ตารางที่ 23 ตารางแสดงผลการเปรียบเทียบประสิทธิภาพวิธีการตรวจหาจุดเปลี่ยนแปลงเมื่อข้อมูลมีขนาด 150

5.1.3 ผลการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 300

ข้อมูลมีการเปลี่ยนแปลงใน	อันดับประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลง
ค่าเฉลี่ย	อันดับที่ 1 วิธี E-Divisive อันดับที่ 2 วิธี e-cp3o อันดับที่ 3 วิธี ks-cp3o
ค่าความแปรปรวน	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี ks-cp3o อันดับที่ 3 วิธี E-Divisive
ค่าสหสัมพันธ์ ($\rho = 0.5$)	อันดับที่ 1 วิธี ks-cp3o อันดับที่ 2 วิธี e-cp3o อันดับที่ 3 วิธี E-Divisive

ข้อมูลมีการเปลี่ยนแปลงใน	อันดับประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลง
ค่าสหสัมพันธ์ ($\rho = 0.9$)	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี ks-cp3o อันดับที่ 3 วิธี E-Divisive
ค่าเฉลี่ย และค่าความแปรปรวน	อันดับที่ 1 วิธี E-Divisive อันดับที่ 2 วิธี e-cp3o อันดับที่ 3 วิธี ks-cp3o
ค่าความแปรปรวน และค่าสหสัมพันธ์	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี ks-cp3o อันดับที่ 3 วิธี E-Divisive
ค่าเฉลี่ย และค่าสหสัมพันธ์	อันดับที่ 1 วิธี E-Divisive อันดับที่ 2 วิธี e-cp3o อันดับที่ 3 วิธี ks-cp3o
ค่าเฉลี่ย, ค่าความแปรปรวน และค่าสหสัมพันธ์	อันดับที่ 1 วิธี E-Divisive อันดับที่ 2 วิธี e-cp3o อันดับที่ 3 วิธี ks-cp3o

ตารางที่ 24 ตารางแสดงผลการเปรียบเทียบประสิทธิภาพวิธีการตรวจหาจุดเปลี่ยนแปลงเมื่อข้อมูลมีขนาด 300

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

5.1.4 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 90

ข้อมูลมีการเปลี่ยนแปลงใน	อันดับความแม่นยำในการตรวจพบจำนวนจุดเปลี่ยนแปลง
ค่าเฉลี่ย	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive
ค่าความแปรปรวน	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive
ค่าสหสัมพันธ์ ($\rho = 0.5$)	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive

ข้อมูลมีการเปลี่ยนแปลงใน	อันดับความแม่นยำในการตรวจพบจำนวนจุดเปลี่ยนแปลง
ค่าสหสัมพันธ์ ($\rho = 0.9$)	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive
ค่าเฉลี่ย และค่าความแปรปรวน	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive
ค่าความแปรปรวน และค่าสหสัมพันธ์	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive
ค่าเฉลี่ย และค่าสหสัมพันธ์	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive
ค่าเฉลี่ย, ค่าความแปรปรวน และค่าสหสัมพันธ์	อันดับที่ 1 วิธี e-cp3o และ ks-cp3o อันดับที่ 2 วิธี E-Divisive

ตารางที่ 25 ตารางแสดงผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อข้อมูลมีขนาด 90

5.1.5 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 150

ข้อมูลมีการเปลี่ยนแปลงใน	อันดับความแม่นยำในการตรวจพบจำนวนจุดเปลี่ยนแปลง
ค่าเฉลี่ย	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี E-Divisive อันดับที่ 3 วิธี ks-cp3o
ค่าความแปรปรวน	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี ks-cp3o อันดับที่ 3 วิธี E-Divisive
ค่าสหสัมพันธ์ ($\rho = 0.5$)	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี ks-cp3o อันดับที่ 3 วิธี E-Divisive
ค่าสหสัมพันธ์ ($\rho = 0.9$)	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี ks-cp3o อันดับที่ 3 วิธี E-Divisive

ข้อมูลมีการเปลี่ยนแปลงใน	อันดับความแม่นยำในการตรวจพบจำนวนจุดเปลี่ยนแปลง
ค่าเฉลี่ย และค่าความแปรปรวน	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี E-Divisive อันดับที่ 3 วิธี ks-cp3o
ค่าความแปรปรวน และค่าสหสัมพันธ์	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี ks-cp3o อันดับที่ 3 วิธี E-Divisive
ค่าเฉลี่ย และค่าสหสัมพันธ์	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี E-Divisive อันดับที่ 3 วิธี ks-cp3o
ค่าเฉลี่ย, ค่าความแปรปรวน และค่าสหสัมพันธ์	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี E-Divisive อันดับที่ 3 วิธี ks-cp3o

ตารางที่ 26 ตารางแสดงผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อข้อมูลมีขนาด 150

5.1.6 ผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อตรวจหาด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อข้อมูลมีขนาดตัวอย่าง 300

ข้อมูลมีการเปลี่ยนแปลงใน	อันดับความแม่นยำในการตรวจพบจำนวนจุดเปลี่ยนแปลง
ค่าเฉลี่ย	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี E-Divisive อันดับที่ 3 วิธี ks-cp3o
ค่าความแปรปรวน	อันดับที่ 1 วิธี E-Divisive อันดับที่ 2 วิธี e-cp3o อันดับที่ 3 วิธี ks-cp3o
ค่าสหสัมพันธ์ ($\rho = 0.5$)	อันดับที่ 1 วิธี E-Divisive อันดับที่ 2 วิธี e-cp3o อันดับที่ 3 วิธี ks-cp3o

ข้อมูลมีการเปลี่ยนแปลงใน	อันดับความแม่นยำในการตรวจพบจำนวนจุดเปลี่ยนแปลง
ค่าสหสัมพันธ์ ($\rho = 0.9$)	อันดับที่ 1 วิธี E-Divisive อันดับที่ 2 วิธี e-cp3o อันดับที่ 3 วิธี ks-cp3o
ค่าเฉลี่ย และค่าความแปรปรวน	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี E-Divisive อันดับที่ 3 วิธี ks-cp3o
ค่าความแปรปรวน และค่าสหสัมพันธ์	อันดับที่ 1 วิธี E-Divisive อันดับที่ 2 วิธี e-cp3o อันดับที่ 3 วิธี ks-cp3o
ค่าเฉลี่ย และค่าสหสัมพันธ์	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี E-Divisive อันดับที่ 3 วิธี ks-cp3o
ค่าเฉลี่ย, ค่าความแปรปรวน และค่าสหสัมพันธ์	อันดับที่ 1 วิธี e-cp3o อันดับที่ 2 วิธี E-Divisive อันดับที่ 3 วิธี ks-cp3o

ตารางที่ 27 ตารางแสดงผลการเปรียบเทียบจำนวนจุดเปลี่ยนแปลงที่พบเมื่อข้อมูลมีขนาด 300

5.1.7 ผลการเปรียบเทียบวิธีการตรวจหาจุดเปลี่ยนแปลงวิธี E-Divisive, e-cp3o และ ks-cp3o เมื่อนำไปตรวจหาจุดเปลี่ยนแปลงในข้อมูลจริง

จากการศึกษาวิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o โดยการนำไปประยุกต์ใช้ในข้อมูลสัญญาณชีพจรคนที่ 2 ของมหาวิทยาลัยควีนส์แลนด์ (The university of Queensland) โดยมีตัวแปรที่สนใจคือ อัตราการเต้นของหัวใจ (Heart rate) และ อัตราการเต้นของชีพจร (Pulse) ซึ่งมีขนาด 900 วินาที (นาทีที่ 0.12 ถึง นาทีที่ 15.11) และข้อมูลปริมาณฝุ่นละออง PM2.5 จาก BERKELEY EARTH ซึ่งมีขนาด 720 จุด (ข้อมูล ณ วันที่ 1 พฤศจิกายน 2561 ถึง 30 พฤศจิกายน 2561) โดยมีตัวแปรที่สนใจคือ ความหนาแน่นของปริมาณฝุ่นละออง PM2.5 และอุณหภูมิโดยทำการตรวจหาข้อมูลทั้งหมดในคราวเดียว และแบ่งข้อมูลออกเป็น 3 ส่วนก่อนแล้วจึงนำแต่ละส่วนมาตรวจหาจุดเปลี่ยนแปลง พบว่า ไม่ว่าจะทำการตรวจหาจากข้อมูล

ทั้งหมดในคราวเดียวกันหรือแบ่งข้อมูลออกเป็น 3 ส่วนก่อน วิธี E-Divisive และ ks-cp3o พบจำนวนจุดและตำแหน่งของจุดเปลี่ยนแปลงที่ใกล้เคียงกัน วิธี e-cp3o พบจำนวนจุดที่เปลี่ยนแปลงน้อยที่สุด

จากการศึกษาเปรียบเทียบจำนวนจุดและตำแหน่งของจุดเปลี่ยนแปลงที่ได้จากการตรวจหาข้อมูลทั้งหมดในคราวเดียว และแบ่งข้อมูลออกเป็น 3 ส่วนก่อนแล้วจึงนำแต่ละส่วนมาตรวจหาจุดเปลี่ยนแปลง พบว่า วิธี E-Divisive และ ks-cp3o มีจำนวนจุดและตำแหน่งของจุดเปลี่ยนแปลงที่ใกล้เคียงกันโดยจำนวนจุดและตำแหน่งของจุดเปลี่ยนแปลงที่ได้จากการแบ่งข้อมูลออกเป็น 3 ส่วนก่อนนั้นมีจำนวนจุดที่ลดลง และตำแหน่งต่างกันเพียงเล็กน้อยเมื่อเทียบกับการตรวจหาข้อมูลทั้งหมดในคราวเดียว และวิธี e-cp3o เมื่อแบ่งข้อมูลออกเป็น 3 ส่วนก่อนนั้นพบจำนวนจุดเปลี่ยนแปลงมากขึ้น โดยมีตำแหน่งเพิ่มขึ้นมาจากการที่ทำการตรวจหาจุดเปลี่ยนแปลงในแต่ละส่วนก่อน ซึ่งหากทำการตรวจหาจุดเปลี่ยนแปลงทั้งหมดในคราวเดียวจะไม่พบจุดเปลี่ยนแปลงนี้

กล่าวคือการตรวจหาจำนวนจุดและตำแหน่งของจุดเปลี่ยนแปลงของข้อมูลสัญญาณชีพและข้อมูลปริมาณฝุ่นละออง PM2.5 โดยใช้วิธีการตรวจหาจุดเปลี่ยนแปลงด้วยวิธี E-Divisive, e-cp3o และ ks-cp3o และพิจารณาความเหมาะสมจากกราฟข้อมูลพบว่า วิธีที่เหมาะสมที่สุดคือ วิธี E-Divisive ซึ่งให้ผลสอดคล้องกับผลที่ได้จากการจำลองข้อมูลนั้นคือ วิธี E-Divisive มีประสิทธิภาพในการตรวจหาจุดเปลี่ยนแปลงมากที่สุดเมื่อข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย

สรุปได้ว่า ภายใต้ขอบเขตการวิจัยในงานวิจัยนี้ เมื่อขนาดตัวอย่างมีขนาดเล็ก วิธีการตรวจหาจุดเปลี่ยนแปลงที่มีประสิทธิภาพมากที่สุดในทุกกรณี คือวิธี e-cp3o และ ks-cp3o และเมื่อขนาดตัวอย่างมีขนาดใหญ่ขึ้น วิธีการตรวจหาจุดเปลี่ยนแปลง E-Divisive มีประสิทธิภาพในการตรวจหาจุดเปลี่ยนแปลงมากที่สุดในกรณีที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย และวิธี e-cp3o มีประสิทธิภาพในการตรวจหาจุดเปลี่ยนแปลงมากที่สุดในกรณีอื่น ๆ ดังนั้นการพิจารณาเลือกใช้วิธีการตรวจหาจุดเปลี่ยนแปลงในกรณีทั่วไปจึงควรเลือกใช้วิธี e-cp3o เนื่องจากวิธี e-cp3o เป็นวิธีการตรวจหาจุดเปลี่ยนแปลงที่มีประสิทธิภาพสูงในทุกกรณี แต่ก็มีข้อจำกัดคือเรื่องของระยะเวลาที่ใช้ในการประมวลผลที่ค่อนข้างนานกว่าวิธี ks-cp3o ที่ให้ประสิทธิภาพใกล้เคียงกับวิธี e-cp3o แต่มีระยะเวลาประมวลผลที่เร็วกว่า และควรเลือกใช้วิธี E-Divisive เมื่อต้องการตรวจหาในกรณีที่ข้อมูลมีการเปลี่ยนแปลงในค่าเฉลี่ย ทั้งนี้การเลือกใช้วิธีการตรวจหาจุดเปลี่ยนแปลงขึ้นอยู่กับวัตถุประสงค์ของงานนั้น ๆ ความยาวของข้อมูล และระยะเวลาที่มีของผู้ที่นำไปใช้งาน

5.2 ข้อเสนอแนะ

1. ในงานวิจัยนี้ทำการเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาจุดเปลี่ยนแปลงเพียง 3 วิธี และเป็นวิธีแบบนอนพาราเมตริกทั้งสิ้น คือวิธี E-Divisive, e-cp3o และ ks-cp3o ผู้วิจัยอื่นอาจทดลองทำการเปรียบเทียบวิธีการตรวจหาจุดเปลี่ยนแปลงวิธีอื่น ๆ ที่น่าสนใจ
2. การจำลองชุดข้อมูลที่นำมาใช้ในการเปรียบเทียบประสิทธิภาพในงานวิจัยนี้ ทำการจำลองเพียงแค่ข้อมูลมีการแจกแจงแบบปกติ 2 ตัวแปร ผู้วิจัยอื่นสามารถเปลี่ยนการแจกแจงของข้อมูลเป็นการแจกแจงอื่น ๆ หรือเปลี่ยนแปลงค่าพารามิเตอร์ หรือเพิ่มขนาดตัวอย่างที่มีขนาดต่างออกไป
3. ข้อมูลที่จำลองขึ้นมา นั้นมีจุดเปลี่ยนแปลงที่เกิดขึ้นเพียง 2 จุด ผู้วิจัยอื่นสามารถเพิ่มจำนวนจุดเปลี่ยนแปลงที่เกิดขึ้นในข้อมูลเพื่อเปรียบเทียบผลของแต่ละวิธีเพิ่มเติมได้
4. ในงานวิจัยนี้ ข้อมูลจำลองที่จำลองขึ้นมา นั้นเป็นข้อมูลอนุกรมเวลาที่มีความผันแปรไม่ปกติ ผู้วิจัยอื่นสามารถทำการจำลองข้อมูลอนุกรมเวลาที่มีค่าแนวโน้ม การเปลี่ยนแปลงตามฤดูกาล หรือการเปลี่ยนแปลงตามวัฏจักร
5. ในส่วนของข้อมูลจริง ในงานวิจัยนี้ผู้วิจัยได้เลือกใช้ตัวแปรที่สนใจเพียงแค่ 2 ตัว ผู้วิจัยอื่นสามารถเลือกใช้ตัวแปรที่สนใจอื่นได้ หรือทำการเพิ่มตัวแปรที่สนใจเพื่อทำการหาจุดเปลี่ยนแปลงได้

บรรณานุกรม

- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1), 193-218. doi:10.1007/bf01908075
- Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500), 1590-1598.
- Liu, D., Görges, M., & Jenkins, S. A. (2012). University of Queensland vital signs dataset: development of an accessible repository of anesthesia patient monitoring data for research. *Anesthesia & Analgesia*, 114(3), 584-589.
- Matteson, D. S., & James, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, 109(505), 334-345.
- Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336), 846-850.
- Szekely, G. J., & Rizzo, M. L. (2005). Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method. *Journal of Classification*, 22(2), 151-183.
- Truong, C., Oudre, L., & Vayatis, N. (2018). A review of change point detection methods. *arXiv preprint arXiv:1808.08751*.

Wang, H., Killick, R., & Fu, X. (2014). Distributional change of monthly precipitation due to climate change: comprehensive examination of dataset in southeastern United States. *Hydrological processes*, 28(20), 5212-5219.

Zhang, W., James, N. A., & Matteson, D. S. (2017). *Pruning and Nonparametric Multiple Change Point Detection*. Paper presented at the 2017 IEEE International Conference on Data Mining Workshops (ICDMW).

วานิชย์บัญชา, ก. (2540). หลักสถิติ: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.



ภาคผนวก

คำสั่งการวิเคราะห์ข้อมูลด้วยโปรแกรม R เวอร์ชัน 3.4.3

```
### Library used in analysis ###
```

```
> library(ecp)
> library(changepoint)
> library(clusteval)
> library(fossil)
> library(clues)
> library(mvtnorm)
> library(devtools)
> #devtools::install_github('alanarnholt/BSDA') for install BSDA
> library(BSDA)
```

```
##### Simulated Data #####
```

```
> ##### case 1 n 90 #####
> base<-c(rep(1,30),rep(2,30),rep(3,30))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(30,mean=c(1,1),sigma=matrix(c(1,0,0,1),ncol=2))
+   period3<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)
+   output2<-e.cp3o(z,K=7)
+   output2cluster<-c()
+   n<-length(output2$estimates)+1
+   for(j in 1:n){
```

```

+   if(j==1){
+     a<-rep(j,output2$estimates[j]-1)
+     output2cluster<-c(output2cluster,a)
+   }
+   if(j!=1&&!is.na(output2$estimates[j])){
+     a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+     output2cluster<-c(output2cluster,a)}
+   if(j!=1&&is.na(output2$estimates[j])){
+     a<-rep(j,90-output2$estimates[j-1]+1)
+     output2cluster<-c(output2cluster,a)
+   }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,90-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1

```

```

+ if(length(cps)==0){
+   peltcluster<-rep(1,90)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,90-cps[j-1]+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> ##### case 2 n 90 #####
> base<-c(rep(1,30),rep(2,30),rep(3,30))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()

```

```

> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(2,0,0,2),ncol=2))
+   period3<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)
+   output2<-e.cp3o(z,K=7)
+   output2cluster<-c()
+   n<-length(output2$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output2$estimates[j]-1)
+       output2cluster<-c(output2cluster,a)
+     }
+     if(j!=1&&!is.na(output2$estimates[j])){
+       a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+       output2cluster<-c(output2cluster,a)}
+     if(j!=1&&is.na(output2$estimates[j])){
+       a<-rep(j,90-output2$estimates[j-1]+1)
+       output2cluster<-c(output2cluster,a)
+     }
+   }
+   output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+   Rand2<-c(Rand2,output2rand)
+   num2<-c(num2,length(output2$estimates))
+   output3<-ks.cp3o(z,K=7)
+   output3cluster<-c()
+   n<-length(output3$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output3$estimates[j]-1)
+       output3cluster<-c(output3cluster,a)
+     }
+     if(j!=1&&!is.na(output3$estimates[j])){
+       a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+       output3cluster<-c(output3cluster,a)}

```

```

+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,90-output3$estimates[j]-1)+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,90)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)}
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,90-cps[j]-1)+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))

```



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

```

> avgRand1
> avgRand2
> avgRand3
> avgRand4

> ##### case 3 n 90 #####
> base<-c(rep(1,30),rep(2,30),rep(3,30))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0.5,0.5,1),ncol=2))
+   period3<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)
+   output2<-e.cp3o(z,K=7)
+   output2cluster<-c()
+   n<-length(output2$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output2$estimates[j]-1)
+       output2cluster<-c(output2cluster,a)
+     }
+     if(j!=1&&!is.na(output2$estimates[j])){
+       a<-rep(j,output2$estimates[j]-output2$estimates[j]-1)
+       output2cluster<-c(output2cluster,a)}
+     if(j!=1&&is.na(output2$estimates[j])){
+       a<-rep(j,90-output2$estimates[j]-1)+1)
+       output2cluster<-c(output2cluster,a)
+     }
  }

```

```

+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,90-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,90)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])

```



```

+   peltcluster<-c(peltcluster,a)
+   if(j!=1&&is.na(cps[j])){
+     a<-rep(j,90-cps[j]-1)+1
+     peltcluster<-c(peltcluster,a)
+   }
+ }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> ##### case 4 n 90 #####
> base<-c(rep(1,30),rep(2,30),rep(3,30))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0.9,0.9,1),ncol=2))
+   period3<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)

```



```

+ output2<-e.cp3o(z,K=7)
+ output2cluster<-c()
+ n<-length(output2$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output2$estimates[j]-1)
+     output2cluster<-c(output2cluster,a)
+   }
+   if(j!=1&&!is.na(output2$estimates[j])){
+     a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+     output2cluster<-c(output2cluster,a)}
+   if(j!=1&&!is.na(output2$estimates[j])){
+     a<-rep(j,90-output2$estimates[j-1]+1)
+     output2cluster<-c(output2cluster,a)
+   }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,90-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")

```

```

+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,90)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)}
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,90-cps[j-1]+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> ##### case 5 n 90 #####
> base<-c(rep(1,30),rep(2,30),rep(3,30))
> Rand1<-c()
> Rand2<-c()
> Rand4<-c()
> Rand3<-c()

```

```

> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+ set.seed(i)
+ period1<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+ period2<-rmvnorm(30,mean=c(1,1),sigma=matrix(c(2,0,0,2),ncol=2))
+ period3<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+ z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+ output1<-e.divisive(z,R=499,alpha=1)
+ output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+ Rand1<-c(Rand1,output1rand)
+ n1<-length(output1$estimates)-2
+ num1<-c(num1,n1)
+ output2<-e.cp3o(z,K=7)
+ output2cluster<-c()
+ n<-length(output2$estimates)+1
+ for(j in 1:n){
+ if(j==1){
+ a<-rep(j,output2$estimates[j])-1
+ output2cluster<-c(output2cluster,a)
+ }
+ if(j!=1&&!is.na(output2$estimates[j])){
+ a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+ output2cluster<-c(output2cluster,a)}
+ if(j!=1&&is.na(output2$estimates[j])){
+ a<-rep(j,90-output2$estimates[j-1]+1)
+ output2cluster<-c(output2cluster,a)
+ }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+ if(j==1){
+ a<-rep(j,output3$estimates[j])-1
+ output3cluster<-c(output3cluster,a)

```

```

+ }
+ if(j!=1&&!is.na(output3$estimates[j])){
+   a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+   output3cluster<-c(output3cluster,a)}
+ if(j!=1&&!is.na(output3$estimates[j])){
+   a<-rep(j,90-output3$estimates[j-1]+1)
+   output3cluster<-c(output3cluster,a)
+ }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,90)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)}
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,90-cps[j-1]+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }

```



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

```

> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> ##### case 6 n 90 #####
> base<-c(rep(1,30),rep(2,30),rep(3,30))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(2,1,1,2),ncol=2))
+   period3<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)
+   output2<-e.cp3o(z,K=7)
+   output2cluster<-c()
+   n<-length(output2$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output2$estimates[j])-1)
+       output2cluster<-c(output2cluster,a)
+     }
+     if(j!=1&&!is.na(output2$estimates[j])){
+       a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+       output2cluster<-c(output2cluster,a)}

```

```

+   if(j!=1&&!is.na(output2$estimates[j])){
+     a<-rep(j,90-output2$estimates[j]-1)+1)
+     output2cluster<-c(output2cluster,a)
+   }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,90-output3$estimates[j]-1)+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,90)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)

```

```

+   peltcluster<-c(peltcluster,a)
+ }
+ if(j!=1&&is.na(cps[j])){
+   a<-rep(j,cps[j]-cps[j-1])
+   peltcluster<-c(peltcluster,a)}
+ if(j!=1&&is.na(cps[j])){
+   a<-rep(j,90-cps[j-1]+1)
+   peltcluster<-c(peltcluster,a)
+ }
+ }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ data<-data.frame(Rand1,Rand2,Rand3,Rand4,num1,num2,num3,num4)
+ colnames(data)<-c('Rand1','Rand2','Rand3','Rand4','num1','num2','num3','num4')
+ write.csv(data,paste('90case6','.csv',sep=' '),row.names=F)
+ print(i)
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> ##### case 7 n 90 #####
> base<-c(rep(1,30),rep(2,30),rep(3,30))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))

```



```

+ period2<-rmvnorm(30,mean=c(1,1),sigma=matrix(c(1,0.5,0.5,1),ncol=2))
+ period3<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+ z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+ output1<-e.divisive(z,R=499,alpha=1)
+ output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+ Rand1<-c(Rand1,output1rand)
+ n1<-length(output1$estimates)-2
+ num1<-c(num1,n1)
+ output2<-e.cp3o(z,K=7)
+ output2cluster<-c()
+ n<-length(output2$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output2$estimates[j]-1)
+     output2cluster<-c(output2cluster,a)
+   }
+   if(j!=1&&!is.na(output2$estimates[j])){
+     a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+     output2cluster<-c(output2cluster,a)}
+   if(j!=1&&is.na(output2$estimates[j])){
+     a<-rep(j,90-output2$estimates[j-1]+1)
+     output2cluster<-c(output2cluster,a)
+   }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,90-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)

```

```

+ }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,90)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)}
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,90-cps[j]-1)+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3

```



```

> avgRand4

> ##### case 8 n 90 #####
> base<-c(rep(1,30),rep(2,30),rep(3,30))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(30,mean=c(1,1),sigma=matrix(c(2,1,1,2),ncol=2))
+   period3<-rmvnorm(30,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)
+   output2<-e.cp3o(z,K=7)
+   output2cluster<-c()
+   n<-length(output2$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output2$estimates[j]-1)
+       output2cluster<-c(output2cluster,a)
+     }
+     if(j!=1&&!is.na(output2$estimates[j])){
+       a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+       output2cluster<-c(output2cluster,a)}
+     if(j!=1&&is.na(output2$estimates[j])){
+       a<-rep(j,90-output2$estimates[j-1]+1)
+       output2cluster<-c(output2cluster,a)
+     }
+   }
+   output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+   Rand2<-c(Rand2,output2rand)

```

```

+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,90-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,90)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)}
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,90-cps[j-1]+1)

```

```

+   peltcluster<-c(peltcluster,a)
+ }
+ }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> #####case 1 n 150 #####
> base<-c(rep(1,50),rep(2,50),rep(3,50))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(50,mean=c(1,1),sigma=matrix(c(1,0,0,1),ncol=2))
+   period3<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)
+   output2<-e.cp3o(z,K=7)
+   output2cluster<-c()
+   n<-length(output2$estimates)+1

```



```

+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output2$estimates[j]-1)
+     output2cluster<-c(output2cluster,a)
+   }
+   if(j!=1&&!is.na(output2$estimates[j])){
+     a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+     output2cluster<-c(output2cluster,a)}
+   if(j!=1&&is.na(output2$estimates[j])){
+     a<-rep(j,150-output2$estimates[j-1]+1)
+     output2cluster<-c(output2cluster,a)
+   }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,150-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()

```

```

+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,150)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)}
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,150-cps[j-1]+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> #####case 2 n 150 #####
> base<-c(rep(1,50),rep(2,50),rep(3,50))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()

```

```

> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(2,0,0,2),ncol=2))
+   period3<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)
+   output2<-e.cp3o(z,K=7)
+   output2cluster<-c()
+   n<-length(output2$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output2$estimates[j]-1)
+       output2cluster<-c(output2cluster,a)
+     }
+     if(j!=1&&!is.na(output2$estimates[j])){
+       a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+       output2cluster<-c(output2cluster,a)}
+     if(j!=1&&is.na(output2$estimates[j])){
+       a<-rep(j,150-output2$estimates[j-1]+1)
+       output2cluster<-c(output2cluster,a)
+     }
+   }
+   output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+   Rand2<-c(Rand2,output2rand)
+   num2<-c(num2,length(output2$estimates))
+   output3<-ks.cp3o(z,K=7)
+   output3cluster<-c()
+   n<-length(output3$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output3$estimates[j]-1)
+       output3cluster<-c(output3cluster,a)
+     }
+     if(j!=1&&!is.na(output3$estimates[j])){
+       a<-rep(j,output3$estimates[j]-output3$estimates[j-1])

```



```

+   output3cluster<-c(output3cluster,a)
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,150-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,150)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,150-cps[j-1]+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))

```

```

> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> #####case 3 n 150 #####
> base<-c(rep(1,50),rep(2,50),rep(3,50))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0.5,0.5,1),ncol=2))
+   period3<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)
+   output2<-e.cp3o(z,K=7)
+   output2cluster<-c()
+   n<-length(output2$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output2$estimates[j]-1)
+       output2cluster<-c(output2cluster,a)
+     }
+     if(j!=1&&!is.na(output2$estimates[j])){
+       a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+       output2cluster<-c(output2cluster,a)}
+     if(j!=1&&is.na(output2$estimates[j])){
+       a<-rep(j,150-output2$estimates[j-1]+1)
+       output2cluster<-c(output2cluster,a)

```

```

+ }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,150-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,150)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){

```

```

+   a<-rep(j,cps[j]-cps[j-1])
+   peltcluster<-c(peltcluster,a)
+   if(j!=1&&is.na(cps[j])){
+     a<-rep(j,150-cps[j-1]+1)
+     peltcluster<-c(peltcluster,a)
+   }
+ }
+ }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> #####case 4 n 150 #####
> base<-c(rep(1,50),rep(2,50),rep(3,50))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0.9,0.9,1),ncol=2))
+   period3<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2

```



```

+ num1<-c(num1,n1)
+ output2<-e.cp3o(z,K=7)
+ output2cluster<-c()
+ n<-length(output2$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output2$estimates[j]-1)
+     output2cluster<-c(output2cluster,a)
+   }
+   if(j!=1&&!is.na(output2$estimates[j])){
+     a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+     output2cluster<-c(output2cluster,a)}
+   if(j!=1&&is.na(output2$estimates[j])){
+     a<-rep(j,150-output2$estimates[j-1]+1)
+     output2cluster<-c(output2cluster,a)
+   }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,150-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])

```

```

+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,150)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,150-cps[j-1]+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> #####case 5 n 150 #####
> base<-c(rep(1,50),rep(2,50),rep(3,50))

```

```

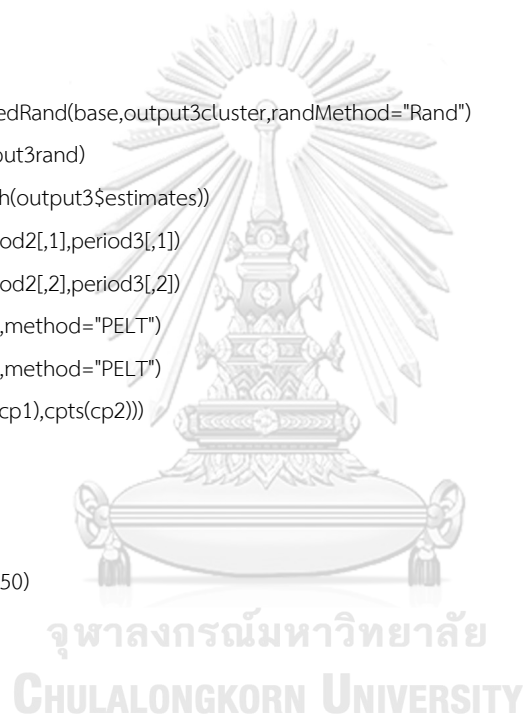
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(50,mean=c(1,1),sigma=matrix(c(2,0,0,2),ncol=2))
+   period3<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)
+   output2<-e.cp3o(z,K=7)
+   output2cluster<-c()
+   n<-length(output2$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output2$estimates[j]-1)
+       output2cluster<-c(output2cluster,a)
+     }
+     if(j!=1&&!is.na(output2$estimates[j])){
+       a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+       output2cluster<-c(output2cluster,a)}
+     if(j!=1&&is.na(output2$estimates[j])){
+       a<-rep(j,150-output2$estimates[j-1]+1)
+       output2cluster<-c(output2cluster,a)
+     }
+   }
+   output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+   Rand2<-c(Rand2,output2rand)
+   num2<-c(num2,length(output2$estimates))
+   output3<-ks.cp3o(z,K=7)
+   output3cluster<-c()
+   n<-length(output3$estimates)+1

```

```

+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,150-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,150)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,150-cps[j-1]+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ }

```




```

+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> #####case 6 n 150 #####
> base<-c(rep(1,50),rep(2,50),rep(3,50))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(2,1,1,2),ncol=2))
+   period3<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)
+   output2<-e.cp3o(z,K=7)
+   output2cluster<-c()
+   n<-length(output2$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output2$estimates[j]-1)
+       output2cluster<-c(output2cluster,a)

```

```

+ }
+ if(j!=1&&!is.na(output2$estimates[j])){
+   a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+   output2cluster<-c(output2cluster,a)}
+ if(j!=1&&is.na(output2$estimates[j])){
+   a<-rep(j,150-output2$estimates[j-1]+1)
+   output2cluster<-c(output2cluster,a)
+ }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,150-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,150)
+ }

```

```

+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)}
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,150-cps[j-1]+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> #####case 7 n 150 #####
> base<-c(rep(1,50),rep(2,50),rep(3,50))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))

```



```

+ period2<-rmvnorm(50,mean=c(1,1),sigma=matrix(c(1,0.5,0.5,1),ncol=2))
+ period3<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+ z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+ output1<-e.divisive(z,R=499,alpha=1)
+ output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+ Rand1<-c(Rand1,output1rand)
+ n1<-length(output1$estimates)-2
+ num1<-c(num1,n1)
+ output2<-e.cp3o(z,K=7)
+ output2cluster<-c()
+ n<-length(output2$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output2$estimates[j]-1)
+     output2cluster<-c(output2cluster,a)
+   }
+   if(j!=1&&!is.na(output2$estimates[j])){
+     a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+     output2cluster<-c(output2cluster,a)}
+   if(j!=1&&is.na(output2$estimates[j])){
+     a<-rep(j,150-output2$estimates[j-1]+1)
+     output2cluster<-c(output2cluster,a)
+   }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,150-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)

```



```

+ }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,150)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)}
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,150-cps[j-1]+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3

```



```

> avgRand4

> #####case 8 n 150 #####
> base<-c(rep(1,50),rep(2,50),rep(3,50))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(50,mean=c(1,1),sigma=matrix(c(2,1,1,2),ncol=2))
+   period3<-rmvnorm(50,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)
+   output2<-e.cp3o(z,K=7)
+   output2cluster<-c()
+   n<-length(output2$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output2$estimates[j]-1)
+       output2cluster<-c(output2cluster,a)
+     }
+     if(j!=1&&!is.na(output2$estimates[j])){
+       a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+       output2cluster<-c(output2cluster,a)}
+     if(j!=1&&is.na(output2$estimates[j])){
+       a<-rep(j,150-output2$estimates[j-1]+1)
+       output2cluster<-c(output2cluster,a)
+     }
+   }
+   output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+   Rand2<-c(Rand2,output2rand)

```

```

+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,150-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,150)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)}
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,150-cps[j-1]+1)

```

```

+   peltcluster<-c(peltcluster,a)
+ }
+ }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> ###case 1 n 300 ###
> base<-c(rep(1,100),rep(2,100),rep(3,100))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num2<-c()
> num1<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(100,mean=c(1,1),sigma=matrix(c(1,0,0,1),ncol=2))
+   period3<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2

```




```

+ num1<-c(num1,n1)
+ output2<-e.cp3o(z,K=7)
+ output2cluster<-c()
+ n<-length(output2$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output2$estimates[j]-1)
+     output2cluster<-c(output2cluster,a)
+   }
+   if(j!=1&&!is.na(output2$estimates[j])){
+     a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+     output2cluster<-c(output2cluster,a)}
+   if(j!=1&&is.na(output2$estimates[j])){
+     a<-rep(j,300-output2$estimates[j-1]+1)
+     output2cluster<-c(output2cluster,a)
+   }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,300-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])

```

```

+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,300)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)}
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,300-cps[j-1]+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> ###case 2 n 300 ###
> base<-c(rep(1,100),rep(2,100),rep(3,100))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()

```

```

> Rand4<-c()
> num2<-c()
> num1<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(2,0,0,2),ncol=2))
+   period3<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)
+   output2<-e.cp3o(z,K=7)
+   output2cluster<-c()
+   n<-length(output2$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output2$estimates[j]-1)
+       output2cluster<-c(output2cluster,a)
+     }
+     if(j!=1&&!is.na(output2$estimates[j])){
+       a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+       output2cluster<-c(output2cluster,a)}
+     if(j!=1&&!is.na(output2$estimates[j])){
+       a<-rep(j,300-output2$estimates[j-1]+1)
+       output2cluster<-c(output2cluster,a)
+     }
+   }
+   output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+   Rand2<-c(Rand2,output2rand)
+   num2<-c(num2,length(output2$estimates))
+   output3<-ks.cp3o(z,K=7)
+   output3cluster<-c()
+   n<-length(output3$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output3$estimates[j]-1)

```

```

+   output3cluster<-c(output3cluster,a)
+ }
+ if(j!=1&&!is.na(output3$estimates[j])){
+   a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+   output3cluster<-c(output3cluster,a)
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,300-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,300)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,300-cps[j-1]+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))

```



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

```

+ }
> avgRand1<-c(mean(Rand1),sd(Rand1))
> avgRand2<-c(mean(Rand2),sd(Rand2))
> avgRand3<-c(mean(Rand3),sd(Rand3))
> avgRand4<-c(mean(Rand4),sd(Rand4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> ###case 3 n 300 ###
> base<-c(rep(1,100),rep(2,100),rep(3,100))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0.5,0.5,1),ncol=2))
+   period3<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)
+   output2<-e.cp3o(z,K=7)
+   output2cluster<-c()
+   n<-length(output2$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output2$estimates[j]-1)
+       output2cluster<-c(output2cluster,a)
+     }
+     if(j!=1&&!is.na(output2$estimates[j])){
+       a<-rep(j,output2$estimates[j]-output2$estimates[j-1])

```

```

+   output2cluster<-c(output2cluster,a)
+   if(j!=1&&is.na(output2$estimates[j])){
+     a<-rep(j,300-output2$estimates[j-1]+1)
+     output2cluster<-c(output2cluster,a)
+   }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,300-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,300)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){

```

```

+   a<-rep(j,cps[j]-1)
+   peltcluster<-c(peltcluster,a)
+ }
+ if(j!=1&&!is.na(cps[j])){
+   a<-rep(j,cps[j]-cps[j-1])
+   peltcluster<-c(peltcluster,a)}
+ if(j!=1&&is.na(cps[j])){
+   a<-rep(j,300-cps[j-1]+1)
+   peltcluster<-c(peltcluster,a)
+ }
+ }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> ###case 4 n 300 ###
> base<-c(rep(1,100),rep(2,100),rep(3,100))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0.9,0.9,1),ncol=2))
+   period3<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)

```

```

+ output1<-e.divisive(z,R=499,alpha=1)
+ output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+ Rand1<-c(Rand1,output1rand)
+ n1<-length(output1$estimates)-2
+ num1<-c(num1,n1)
+ output2<-e.cp3o(z,K=7)
+ output2cluster<-c()
+ n<-length(output2$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output2$estimates[j]-1)
+     output2cluster<-c(output2cluster,a)
+   }
+   if(j!=1&&!is.na(output2$estimates[j])){
+     a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+     output2cluster<-c(output2cluster,a)}
+   if(j!=1&&is.na(output2$estimates[j])){
+     a<-rep(j,300-output2$estimates[j-1]+1)
+     output2cluster<-c(output2cluster,a)
+   }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,300-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")

```



```

+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,300)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,300-cps[j-1]+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> ###case 5 n 300 ###

```

```

> base<-c(rep(1,100),rep(2,100),rep(3,100))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(100,mean=c(1,1),sigma=matrix(c(2,0,0,2),ncol=2))
+   period3<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)
+   output2<-e.cp3o(z,K=7)
+   output2cluster<-c()
+   n<-length(output2$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output2$estimates[j]-1)
+       output2cluster<-c(output2cluster,a)
+     }
+     if(j!=1&&!is.na(output2$estimates[j])){
+       a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+       output2cluster<-c(output2cluster,a)}
+     if(j!=1&&is.na(output2$estimates[j])){
+       a<-rep(j,300-output2$estimates[j-1]+1)
+       output2cluster<-c(output2cluster,a)
+     }
+   }
+   output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+   Rand2<-c(Rand2,output2rand)
+   num2<-c(num2,length(output2$estimates))
+   output3<-ks.cp3o(z,K=7)
+   output3cluster<-c()

```

```

+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,300-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,300)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)}
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,300-cps[j-1]+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }

```

```

+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

> ###case 6 n 300 ###
> base<-c(rep(1,100),rep(2,100),rep(3,100))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)
+   period1<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   period2<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(2,1,1,2),ncol=2))
+   period3<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+   z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+   output1<-e.divisive(z,R=499,alpha=1)
+   output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+   Rand1<-c(Rand1,output1rand)
+   n1<-length(output1$estimates)-2
+   num1<-c(num1,n1)
+   output2<-e.cp3o(z,K=7)
+   output2cluster<-c()
+   n<-length(output2$estimates)+1
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,output2$estimates[j]-1)

```



```

+   output2cluster<-c(output2cluster,a)
+ }
+ if(j!=1&&!is.na(output2$estimates[j])){
+   a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+   output2cluster<-c(output2cluster,a)
+   if(j!=1&&!is.na(output2$estimates[j])){
+     a<-rep(j,300-output2$estimates[j-1]+1)
+     output2cluster<-c(output2cluster,a)
+   }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,300-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,300)

```

```

+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)}
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,300-cps[j-1]+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

```

```

> ###case 7 n 300 ###
> base<-c(rep(1,100),rep(2,100),rep(3,100))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+   set.seed(i)

```



```

+ period1<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+ period2<-rmvnorm(100,mean=c(1,1),sigma=matrix(c(1,0.5,0.5,1),ncol=2))
+ period3<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+ z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+ output1<-e.divisive(z,R=499,alpha=1)
+ output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+ Rand1<-c(Rand1,output1rand)
+ n1<-length(output1$estimates)-2
+ num1<-c(num1,n1)
+ output2<-e.cp3o(z,K=7)
+ output2cluster<-c()
+ n<-length(output2$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output2$estimates[j]-1)
+     output2cluster<-c(output2cluster,a)
+   }
+   if(j!=1&&!is.na(output2$estimates[j])){
+     a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+     output2cluster<-c(output2cluster,a)}
+   if(j!=1&&is.na(output2$estimates[j])){
+     a<-rep(j,300-output2$estimates[j-1]+1)
+     output2cluster<-c(output2cluster,a)
+   }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")
+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,300-output3$estimates[j-1]+1)

```

```

+   output3cluster<-c(output3cluster,a)
+ }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,300)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)}
+     if(j!=1&&is.na(cps[j])){
+       a<-rep(j,300-cps[j-1]+1)
+       peltcluster<-c(peltcluster,a)
+     }
+   }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2

```



```

> avgRand3
> avgRand4

> ###case 8 n 300 ###
> base<-c(rep(1,100),rep(2,100),rep(3,100))
> Rand1<-c()
> Rand2<-c()
> Rand3<-c()
> Rand4<-c()
> num1<-c()
> num2<-c()
> num3<-c()
> num4<-c()
> for(i in 1:500){
+ set.seed(i)
+ period1<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+ period2<-rmvnorm(100,mean=c(1,1),sigma=matrix(c(2,1,1,2),ncol=2))
+ period3<-rmvnorm(100,mean=c(0,0),sigma=matrix(c(1,0,0,1),ncol=2))
+ z<-matrix(c(period1[,1],period2[,1],period3[,1],period1[,2],period2[,2],period3[,2]),ncol=2)
+ output1<-e.divisive(z,R=499,alpha=1)
+ output1rand<-adjustedRand(base,output1$cluster,randMethod = "Rand")
+ Rand1<-c(Rand1,output1rand)
+ n1<-length(output1$estimates)-2
+ num1<-c(num1,n1)
+ output2<-e.cp3o(z,K=7)
+ output2cluster<-c()
+ n<-length(output2$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output2$estimates[j]-1)
+     output2cluster<-c(output2cluster,a)
+   }
+   if(j!=1&&!is.na(output2$estimates[j])){
+     a<-rep(j,output2$estimates[j]-output2$estimates[j-1])
+     output2cluster<-c(output2cluster,a)}
+   if(j!=1&&is.na(output2$estimates[j])){
+     a<-rep(j,300-output2$estimates[j-1]+1)
+     output2cluster<-c(output2cluster,a)
+   }
+ }
+ output2rand<-adjustedRand(base,output2cluster,randMethod="Rand")

```

```

+ Rand2<-c(Rand2,output2rand)
+ num2<-c(num2,length(output2$estimates))
+ output3<-ks.cp3o(z,K=7)
+ output3cluster<-c()
+ n<-length(output3$estimates)+1
+ for(j in 1:n){
+   if(j==1){
+     a<-rep(j,output3$estimates[j]-1)
+     output3cluster<-c(output3cluster,a)
+   }
+   if(j!=1&&!is.na(output3$estimates[j])){
+     a<-rep(j,output3$estimates[j]-output3$estimates[j-1])
+     output3cluster<-c(output3cluster,a)}
+   if(j!=1&&is.na(output3$estimates[j])){
+     a<-rep(j,300-output3$estimates[j-1]+1)
+     output3cluster<-c(output3cluster,a)
+   }
+ }
+ output3rand<-adjustedRand(base,output3cluster,randMethod="Rand")
+ Rand3<-c(Rand3,output3rand)
+ num3<-c(num3,length(output3$estimates))
+ v1<-c(period1[,1],period2[,1],period3[,1])
+ v2<-c(period1[,2],period2[,2],period3[,2])
+ cp1<-cpt.meanvar(v1,method="PELT")
+ cp2<-cpt.meanvar(v2,method="PELT")
+ cps<-sort(union(cpts(cp1),cpts(cp2)))
+ peltcluster<-c()
+ n<-length(cps)+1
+ if(length(cps)==0){
+   peltcluster<-rep(1,300)
+ }
+ if(length(cps)!=0){
+   for(j in 1:n){
+     if(j==1){
+       a<-rep(j,cps[j]-1)
+       peltcluster<-c(peltcluster,a)
+     }
+     if(j!=1&&!is.na(cps[j])){
+       a<-rep(j,cps[j]-cps[j-1])
+       peltcluster<-c(peltcluster,a)}
+     if(j!=1&&is.na(cps[j])){

```

```

+   a<-rep(j,300-cps[j-1]+1)
+   peltcluster<-c(peltcluster,a)
+ }
+ }
+ }
+ peltrand<-adjustedRand(base,peltcluster,randMethod = "Rand")
+ Rand4<-c(Rand4,peltrand)
+ num4<-c(num4,length(cps))
+ }
> avgRand1<-c(mean(Rand1),sd(Rand1),mean(num1))
> avgRand2<-c(mean(Rand2),sd(Rand2),mean(num2))
> avgRand3<-c(mean(Rand3),sd(Rand3),mean(num3))
> avgRand4<-c(mean(Rand4),sd(Rand4),mean(num4))
> avgRand1
> avgRand2
> avgRand3
> avgRand4

##### Real Data
#####

##### Vital Sign Data #####
### Full Data ###
> data<-read.csv('realdata02.csv',sep=',',header=T)
> hr<-data$HR
> pulse<-data$Pulse
> z<-matrix(c(hr,pulse),ncol=2)
> colnames(z)<-c("HR","Pulse")
> zhr<-z[,1]
> zp<-z[,2]
> output1<-e.divisive(z,R=499,alpha=1)
> cps1<-output1$estimates[-c(1,length(output1$estimates))]
> MeanHR1<-c()
> MeanP1<-c()
> VarHR1<-c()
> VarP1<-c()
> Corr1<-c()
> for(i in 0:length(cps1)){
+   if(i==0){
+     MeanHR1<-c(MeanHR1,mean(zhr[1:(cps1[i+1]-1)]))
+     MeanP1<-c(MeanP1,mean(zp[1:(cps1[i+1]-1)]))

```

```

+ VarHR1<-c(VarHR1,var(zhr[1:(cps1[i+1]-1)]))
+ VarP1<-c(VarP1,var(zp[1:(cps1[i+1]-1)]))
+ Corr1<-c(Corr1,cor(z[1:(cps1[i+1]-1),][2]))
+ }
+ if(i>0&&i<length(cps1)){
+ MeanHR1<-c(MeanHR1,mean(zhr[cps1[i]:(cps1[i+1]-1)]))
+ MeanP1<-c(MeanP1,mean(zp[cps1[i]:(cps1[i+1]-1)]))
+ VarHR1<-c(VarHR1,var(zhr[cps1[i]:(cps1[i+1]-1)]))
+ VarP1<-c(VarP1,var(zp[cps1[i]:(cps1[i+1]-1)]))
+ Corr1<-c(Corr1,cor(z[cps1[i]:(cps1[i+1]-1),][2]))
+ }
+ if(i==length(cps1)){
+ MeanHR1<-c(MeanHR1,mean(zhr[cps1[i]:900]))
+ MeanP1<-c(MeanP1,mean(zp[cps1[i]:900]))
+ VarHR1<-c(VarHR1,var(zhr[cps1[i]:900]))
+ VarP1<-c(VarP1,var(zp[cps1[i]:900]))
+ Corr1<-c(Corr1,cor(z[cps1[i]:900,][2]))
+ }
+ }
> MeanHR1
> MeanP1
> VarHR1
> VarP1
> Corr1
> hrz1<-c()
> hrf1<-c()
> pz1<-c()
> pf1<-c()
> for(i in 0:(length(cps1)-1)){
+ if(i==0){
+ x1<-zhr[1:(cps1[i+1]-1)]
+ y1<-zhr[cps1[i+1]:(cps1[i+2]-1)]
+ x2<-zp[1:(cps1[i+1]-1)]
+ y2<-zp[cps1[i+1]:(cps1[i+2]-1)]
+ z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+ f1<-var.test(x1,y1)
+ z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+ f2<-var.test(x2,y2)
+ hrz1<-c(hrz1,round(z1$p.value,4))
+ hrf1<-c(hrf1,round(f1$p.value,4))
+ pz1<-c(pz1,round(z2$p.value,4))

```



```

+ pf1<-c(pf1,round(f2$p.value,4))
+ }
+ if(i>0&&!(length(cps1)-1)){
+ x1<-zhr[cps1[i]:(cps1[i+1]-1)]
+ y1<-zhr[cps1[i+1]:(cps1[i+2]-1)]
+ x2<-zp[cps1[i]:(cps1[i+1]-1)]
+ y2<-zp[cps1[i+1]:(cps1[i+2]-1)]
+ z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+ f1<-var.test(x1,y1)
+ z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+ f2<-var.test(x2,y2)
+ hrz1<-c(hrz1,round(z1$p.value,4))
+ hrf1<-c(hrf1,round(f1$p.value,4))
+ pz1<-c(pz1,round(z2$p.value,4))
+ pf1<-c(pf1,round(f2$p.value,4))
+ }
+ if(i==length(cps1)-1){
+ x1<-zhr[cps1[i]:(cps1[i+1]-1)]
+ y1<-zhr[cps1[i+1]:900]
+ x2<-zp[cps1[i]:(cps1[i+1]-1)]
+ y2<-zp[cps1[i+1]:900]
+ z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+ f1<-var.test(x1,y1)
+ z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+ f2<-var.test(x2,y2)
+ hrz1<-c(hrz1,round(z1$p.value,4))
+ hrf1<-c(hrf1,round(f1$p.value,4))
+ pz1<-c(pz1,round(z2$p.value,4))
+ pf1<-c(pf1,round(f2$p.value,4))
+ }
+ }
> hrz1
> hrf1
> pz1
> pf1

> output2<-e.cp3o(z,K=300)
> cps2<-output2$estimates
> MeanHR2<-c()
> MeanP2<-c()

```

```

> VarHR2<-c()
> VarP2<-c()
> Corr2<-c()
> for(i in 0:length(cps2)){
+   if(i==0){
+     MeanHR2<-c(MeanHR2,mean(zhr[1:(cps2[i+1]-1)]))
+     MeanP2<-c(MeanP2,mean(zp[1:(cps2[i+1]-1)]))
+     VarHR2<-c(VarHR2,var(zhr[1:(cps2[i+1]-1)]))
+     VarP2<-c(VarP2,var(zp[1:(cps2[i+1]-1)]))
+     Corr2<-c(Corr2,cor(z[1:(cps2[i+1]-1)],)[2])
+   }
+   if(i>0&&i<length(cps2)){
+     MeanHR2<-c(MeanHR2,mean(zhr[cps2[i]:(cps2[i+1]-1)]))
+     MeanP2<-c(MeanP2,mean(zp[cps2[i]:(cps2[i+1]-1)]))
+     VarHR2<-c(VarHR2,var(zhr[cps2[i]:(cps2[i+1]-1)]))
+     VarP2<-c(VarP2,var(zp[cps2[i]:(cps2[i+1]-1)]))
+     Corr2<-c(Corr2,cor(z[cps2[i]:(cps2[i+1]-1)],)[2])
+   }
+   if(i==length(cps2)){
+     MeanHR2<-c(MeanHR2,mean(zhr[cps2[i]:900]))
+     MeanP2<-c(MeanP2,mean(zp[cps2[i]:900]))
+     VarHR2<-c(VarHR2,var(zhr[cps2[i]:900]))
+     VarP2<-c(VarP2,var(zp[cps2[i]:900]))
+     Corr2<-c(Corr2,cor(z[cps2[i]:900,])[2])
+   }
+ }
> MeanHR2
> MeanP2
> VarHR2
> VarP2
> Corr2
> hrz2<-c()
> hrf2<-c()
> pzf2<-c()
> pf2<-c()
> for(i in 0:(length(cps2)-1)){
+   if(i==0){
+     x1<-zhr[1:(cps2[i+1]-1)]
+     y1<-zhr[cps2[i+1]:(cps2[i+2]-1)]
+     x2<-zp[1:(cps2[i+1]-1)]
+     y2<-zp[cps2[i+1]:(cps2[i+2]-1)]

```



```

+ z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+ f1<-var.test(x1,y1)
+ z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+ f2<-var.test(x2,y2)
+ hrz2<-c(hrz2,round(z1$p.value,4))
+ hrf2<-c(hrf2,round(f1$p.value,4))
+ pz2<-c(pz2,round(z2$p.value,4))
+ pf2<-c(pf2,round(f2$p.value,4))
+ }
+ if(i>0&&i<(length(cps2)-1)){
+ x1<-zhr[cps2[i]:cps2[i+1]-1]
+ y1<-zhr[cps2[i+1]:cps2[i+2]-1]
+ x2<-zp[cps2[i]:cps2[i+1]-1]
+ y2<-zp[cps2[i+1]:cps2[i+2]-1]
+ z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+ f1<-var.test(x1,y1)
+ z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+ f2<-var.test(x2,y2)
+ hrz2<-c(hrz2,round(z1$p.value,4))
+ hrf2<-c(hrf2,round(f1$p.value,4))
+ pz2<-c(pz2,round(z2$p.value,4))
+ pf2<-c(pf2,round(f2$p.value,4))
+ }
+ if(i==length(cps2)-1){
+ x1<-zhr[cps2[i]:cps2[i+1]-1]
+ y1<-zhr[cps2[i+1]:900]
+ x2<-zp[cps2[i]:cps2[i+1]-1]
+ y2<-zp[cps2[i+1]:900]
+ z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+ f1<-var.test(x1,y1)
+ z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+ f2<-var.test(x2,y2)
+ hrz2<-c(hrz2,round(z1$p.value,4))
+ hrf2<-c(hrf2,round(f1$p.value,4))
+ pz2<-c(pz2,round(z2$p.value,4))
+ pf2<-c(pf2,round(f2$p.value,4))
+ }
+ }
> hrz2
> hrf2
> pz2

```

```

> pf2

> output3<-ks.cp3o(z,K=300)
> cps3<-output3$estimates
> MeanHR3<-c()
> MeanP3<-c()
> VarHR3<-c()
> VarP3<-c()
> Corr3<-c()
> for(i in 0:length(cps3)){
+   if(i==0){
+     MeanHR3<-c(MeanHR3,mean(zhr[1:(cps3[i+1]-1)]))
+     MeanP3<-c(MeanP3,mean(zp[1:(cps3[i+1]-1)]))
+     VarHR3<-c(VarHR3,var(zhr[1:(cps3[i+1]-1)]))
+     VarP3<-c(VarP3,var(zp[1:(cps3[i+1]-1)]))
+     Corr3<-c(Corr3,cor(z[1:(cps3[i+1]-1),,][2]))
+   }
+   if(i>0&&i<length(cps3)){
+     MeanHR3<-c(MeanHR3,mean(zhr[cps3[i]:cps3[i+1]-1]))
+     MeanP3<-c(MeanP3,mean(zp[cps3[i]:cps3[i+1]-1]))
+     VarHR3<-c(VarHR3,var(zhr[cps3[i]:cps3[i+1]-1]))
+     VarP3<-c(VarP3,var(zp[cps3[i]:cps3[i+1]-1]))
+     Corr3<-c(Corr3,cor(z[cps3[i]:cps3[i+1]-1,][2]))
+   }
+   if(i==length(cps3)){
+     MeanHR3<-c(MeanHR3,mean(zhr[cps3[i]:900]))
+     MeanP3<-c(MeanP3,mean(zp[cps3[i]:900]))
+     VarHR3<-c(VarHR3,var(zhr[cps3[i]:900]))
+     VarP3<-c(VarP3,var(zp[cps3[i]:900]))
+     Corr3<-c(Corr3,cor(z[cps3[i]:900,][2]))
+   }
+ }
> MeanHR3
> MeanP3
> VarHR3
> VarP3
> Corr3
> hrz3<-c()
> hrf3<-c()
> pz3<-c()

```



```

> pf3<-c()
> for(i in 0:(length(cps3)-1)){
+   if(i==0){
+     x1<-zhr[1:(cps3[i+1]-1)]
+     y1<-zhr[cps3[i+1]:cps3[i+2]-1]]
+     x2<-zp[1:(cps3[i+1]-1)]
+     y2<-zp[cps3[i+1]:cps3[i+2]-1]]
+     z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+     f1<-var.test(x1,y1)
+     z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+     f2<-var.test(x2,y2)
+     hrz3<-c(hrz3,round(z1$p.value,4))
+     hrf3<-c(hrf3,round(f1$p.value,4))
+     pz3<-c(pz3,round(z2$p.value,4))
+     pf3<-c(pf3,round(f2$p.value,4))
+   }
+   if(i>0&&i<(length(cps3)-1)){
+     x1<-zhr[cps3[i]:cps3[i+1]-1]]
+     y1<-zhr[cps3[i+1]:cps3[i+2]-1]]
+     x2<-zp[cps3[i]:cps3[i+1]-1]]
+     y2<-zp[cps3[i+1]:cps3[i+2]-1]]
+     z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+     f1<-var.test(x1,y1)
+     z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+     f2<-var.test(x2,y2)
+     hrz3<-c(hrz3,round(z1$p.value,4))
+     hrf3<-c(hrf3,round(f1$p.value,4))
+     pz3<-c(pz3,round(z2$p.value,4))
+     pf3<-c(pf3,round(f2$p.value,4))
+   }
+   if(i==length(cps3)-1){
+     x1<-zhr[cps3[i]:cps3[i+1]-1]]
+     y1<-zhr[cps3[i+1]:900]
+     x2<-zp[cps3[i]:cps3[i+1]-1]]
+     y2<-zp[cps3[i+1]:900]
+     z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+     f1<-var.test(x1,y1)
+     z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+     f2<-var.test(x2,y2)
+     hrz3<-c(hrz3,round(z1$p.value,4))
+     hrf3<-c(hrf3,round(f1$p.value,4))

```

```

+ pz3<-c(pz3,round(z2$p.value,4))
+ pf3<-c(pf3,round(f2$p.value,4))
+ }
+ }
> hrz3
> hrf3
> pz3
> pf3

```

```
### Split Data ###
```

```

> data<-read.csv('realdata02.csv',sep=',',header=T)
> hr<-data$HR
> pulse<-data$Pulse
> z<-matrix(c(hr,pulse),ncol=2)
> colnames(z)<-c("HR","Pulse")
> z1<-matrix(c(hr[1:300],pulse[1:300]),ncol=2)
> z2<-matrix(c(hr[301:600],pulse[301:600]),ncol=2)
> z3<-matrix(c(hr[601:900],pulse[601:900]),ncol=2)
> output11<-e.divisive(z1,R=499,alpha=1)
> output12<-e.divisive(z2,R=499,alpha=1)
> output13<-e.divisive(z3,R=499,alpha=1)
> output21<-e.cp3o(z1,K=300)
> output22<-e.cp3o(z2,K=300)
> output23<-e.cp3o(z3,K=300)
> output31<-ks.cp3o(z1,K=300)
> output32<-ks.cp3o(z2,K=300)
> output33<-ks.cp3o(z3,K=300)
> output4<-c()
> for(i in 1:3){
+   a<-300*(i-1)
+   cp1<-cpt.meanvar(z[,1][1:(1+a):(300+a)],method="PELT")
+   cp2<-cpt.meanvar(z[,2][1:(1+a):(300+a)],method="PELT")
+   cps<-sort(union(cpts(cp1),cpts(cp2)))
+   if(i==2||i==3){
+     cps<-sort(union(cpts(cp1),cpts(cp2))+(300*(i-1)))
+   }
+   output4<-c(output4,cps)
+ }
> output1<-c(output11$estimates[-c(1,length(output11$estimates))],output12$estimates[-
c(1,length(output12$estimates))]+300,output13$estimates[-c(1,length(output13$estimates))]+600)

```

```

> output2<-c(output21$estimates,output22$estimates+300,output23$estimates+600)
> output3<-c(output31$estimates,output32$estimates+300,output33$estimates+600)
> output1
> output2
> output3
> output4

```

```
##### PM2.5 Data #####
```

```
##### Full Data #####
```

```

> data<-read.csv('dust.csv',sep=',',header=T)
> pm<-data$PM
> temp<-data$Temp
> z<-matrix(c(pm,temp),ncol=2)
> colnames(z)<-c("PM2.5","Temperature")
> zpm<-z[,1]
> zt<-z[,2]
> output1<-e.divisive(z,R=499,alpha=1)
> cps1<-output1$estimates[-c(1,length(output1$estimates))]
> MeanPM1<-c()
> MeanT1<-c()
> VarPM1<-c()
> VarT1<-c()
> Corr1<-c()
> for(i in 0:length(cps1)){
+   if(i==0){
+     MeanPM1<-c(MeanPM1,mean(zpm[1:(cps1[i+1]-1)]))
+     MeanT1<-c(MeanT1,mean(zt[1:(cps1[i+1]-1)]))
+     VarPM1<-c(VarPM1,var(zpm[1:(cps1[i+1]-1)]))
+     VarT1<-c(VarT1,var(zt[1:(cps1[i+1]-1)]))
+     Corr1<-c(Corr1,cor(z[1:(cps1[i+1]-1),,][2])
+   }
+   if(i>0&&i<length(cps1)){
+     MeanPM1<-c(MeanPM1,mean(zpm[cps1[i]:(cps1[i+1]-1)]))
+     MeanT1<-c(MeanT1,mean(zt[cps1[i]:(cps1[i+1]-1)]))
+     VarPM1<-c(VarPM1,var(zpm[cps1[i]:(cps1[i+1]-1)]))
+     VarT1<-c(VarT1,var(zt[cps1[i]:(cps1[i+1]-1)]))
+     Corr1<-c(Corr1,cor(z[cps1[i]:(cps1[i+1]-1),,][2])
+   }
+   if(i==length(cps1)){
+     MeanPM1<-c(MeanPM1,mean(zpm[cps1[i]:720]))
+     MeanT1<-c(MeanT1,mean(zt[cps1[i]:720]))

```

```

+ VarPM1<-c(VarPM1,var(zpm[cps1[i]:720]))
+ VarT1<-c(VarT1,var(zt[cps1[i]:720]))
+ Corr1<-c(Corr1,cor(z[cps1[i]:720,])[2])
+ }
+ }
> MeanPM1
> MeanT1
> VarPM1
> VarT1
> Corr1
> pmz1<-c()
> pmf1<-c()
> tz1<-c()
> tf1<-c()
> for(i in 0:(length(cps1)-1)){
+   if(i==0){
+     x1<-zpm[1:(cps1[i+1]-1)]
+     y1<-zpm[cps1[i+1]:(cps1[i+2]-1)]
+     x2<-zt[1:(cps1[i+1]-1)]
+     y2<-zt[cps1[i+1]:(cps1[i+2]-1)]
+     z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+     f1<-var.test(x1,y1)
+     z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+     f2<-var.test(x2,y2)
+     pmz1<-c(pmz1,round(z1$p.value,4))
+     pmf1<-c(pmf1,round(f1$p.value,4))
+     tz1<-c(tz1,round(z2$p.value,4))
+     tf1<-c(tf1,round(f2$p.value,4))
+   }
+   if(i>0&&i<(length(cps1)-1)){
+     x1<-zpm[cps1[i]:(cps1[i+1]-1)]
+     y1<-zpm[cps1[i+1]:(cps1[i+2]-1)]
+     x2<-zt[cps1[i]:(cps1[i+1]-1)]
+     y2<-zt[cps1[i+1]:(cps1[i+2]-1)]
+     z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+     f1<-var.test(x1,y1)
+     z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+     f2<-var.test(x2,y2)
+     pmz1<-c(pmz1,round(z1$p.value,4))
+     pmf1<-c(pmf1,round(f1$p.value,4))
+     tz1<-c(tz1,round(z2$p.value,4))

```

```

+ tf1<-c(tf1,round(f2$p.value,4))
+ }
+ if(i==length(cps1)-1){
+ x1<-zpm[cps1[i]:(cps1[i+1]-1)]
+ y1<-zpm[cps1[i+1]:720]
+ x2<-zt[cps1[i]:(cps1[i+1]-1)]
+ y2<-zt[cps1[i+1]:720]
+ z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+ f1<-var.test(x1,y1)
+ z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+ f2<-var.test(x2,y2)
+ pmz1<-c(pmz1,round(z1$p.value,4))
+ pmf1<-c(pmf1,round(f1$p.value,4))
+ tz1<-c(tz1,round(z2$p.value,4))
+ tf1<-c(tf1,round(f2$p.value,4))
+ }
+ }
> pmz1
> pmf1
> tz1
> tf1

> output2<-e.cp3o(z,K=300)
> cps2<-output2$estimates
> MeanPM2<-c()
> MeanT2<-c()
> VarPM2<-c()
> VarT2<-c()
> Corr2<-c()
> for(i in 0:length(cps2)){
+ if(i==0){
+ MeanPM2<-c(MeanPM2,mean(zpm[1:(cps2[i+1]-1)]))
+ MeanT2<-c(MeanT2,mean(zt[1:(cps2[i+1]-1)]))
+ VarPM2<-c(VarPM2,var(zpm[1:(cps2[i+1]-1)]))
+ VarT2<-c(VarT2,var(zt[1:(cps2[i+1]-1)]))
+ Corr2<-c(Corr2,cor(z[1:(cps2[i+1]-1),,][2])
+ }
+ if(i>0&&i<length(cps2)){
+ MeanPM2<-c(MeanPM2,mean(zpm[cps2[i]:(cps2[i+1]-1)]))
+ MeanT2<-c(MeanT2,mean(zt[cps2[i]:(cps2[i+1]-1)]))
+ VarPM2<-c(VarPM2,var(zpm[cps2[i]:(cps2[i+1]-1)]))

```



```

+ VarT2<-c(VarT2,var(zt[cps2[i]:(cps2[i+1]-1)]))
+ Corr2<-c(Corr2,cor(z[cps2[i]:(cps2[i+1]-1)],)[2])
+ }
+ if(i==length(cps2)){
+ MeanPM2<-c(MeanPM2,mean(zpm[cps2[i]:720]))
+ MeanT2<-c(MeanT2,mean(zt[cps2[i]:720]))
+ VarPM2<-c(VarPM2,var(zpm[cps2[i]:720]))
+ VarT2<-c(VarT2,var(zt[cps2[i]:720]))
+ Corr2<-c(Corr2,cor(z[cps2[i]:720,])[2])
+ }
+ }
> MeanPM2
> MeanT2
> VarPM2
> VarT2
> Corr2
> pmz2<-c()
> pmf2<-c()
> tz2<-c()
> tf2<-c()
> for(i in 0:(length(cps2)-1)){
+ if(i==0){
+ x1<-zpm[1:(cps2[i+1]-1)]
+ y1<-zpm[cps2[i+1]:(cps2[i+2]-1)]
+ x2<-zt[1:(cps2[i+1]-1)]
+ y2<-zt[cps2[i+1]:(cps2[i+2]-1)]
+ z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+ f1<-var.test(x1,y1)
+ z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+ f2<-var.test(x2,y2)
+ pmz2<-c(pmz2,round(z1$p.value,4))
+ pmf2<-c(pmf2,round(f1$p.value,4))
+ tz2<-c(tz2,round(z2$p.value,4))
+ tf2<-c(tf2,round(f2$p.value,4))
+ }
+ if(i>0&&i<(length(cps2)-1)){
+ x1<-zpm[cps2[i]:(cps2[i+1]-1)]
+ y1<-zpm[cps2[i+1]:(cps2[i+2]-1)]
+ x2<-zt[cps2[i]:(cps2[i+1]-1)]
+ y2<-zt[cps2[i+1]:(cps2[i+2]-1)]
+ z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))

```



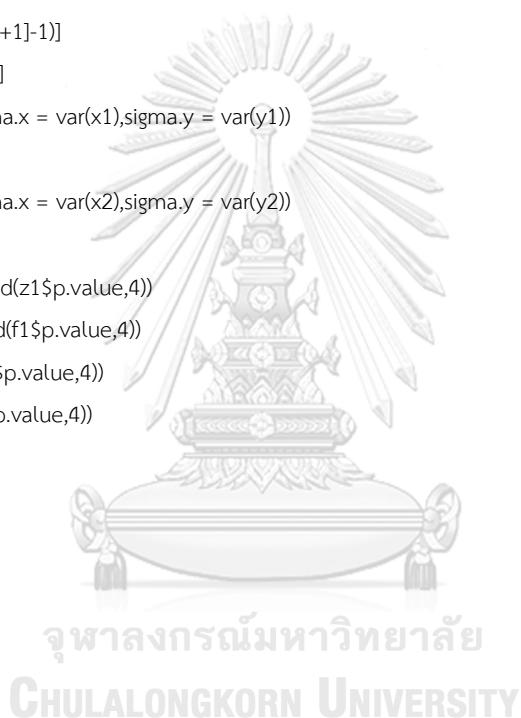
วิทยาลัย
 CHULALONGKORN UNIVERSITY

```

+ f1<-var.test(x1,y1)
+ z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+ f2<-var.test(x2,y2)
+ pmz2<-c(pmz2,round(z1$p.value,4))
+ pmf2<-c(pmf2,round(f1$p.value,4))
+ tz2<-c(tz2,round(z2$p.value,4))
+ tf2<-c(tf2,round(f2$p.value,4))
+ }
+ if(i==length(cps2)-1){
+ x1<-zpm[cps2[i]:cps2[i+1]-1]
+ y1<-zpm[cps2[i+1]:720]
+ x2<-zt[cps2[i]:cps2[i+1]-1]
+ y2<-zt[cps2[i+1]:720]
+ z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+ f1<-var.test(x1,y1)
+ z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+ f2<-var.test(x2,y2)
+ pmz2<-c(pmz2,round(z1$p.value,4))
+ pmf2<-c(pmf2,round(f1$p.value,4))
+ tz2<-c(tz2,round(z2$p.value,4))
+ tf2<-c(tf2,round(f2$p.value,4))
+ }
+ }
> pmz2
> pmf2
> tz2
> tf2

> output3<-ks.cp3o(z,K=300)
> cps3<-output3$estimates
> MeanPM3<-c()
> MeanT3<-c()
> VarPM3<-c()
> VarT3<-c()
> Corr3<-c()
> for(i in 0:length(cps3)){
+ if(i==0){
+ MeanPM3<-c(MeanPM3,mean(zpm[1:(cps3[i+1]-1)]))
+ MeanT3<-c(MeanT3,mean(zt[1:(cps3[i+1]-1)]))
+ VarPM3<-c(VarPM3,var(zpm[1:(cps3[i+1]-1)]))
+ VarT3<-c(VarT3,var(zt[1:(cps3[i+1]-1)]))

```



```

+ Corr3<-c(Corr3,cor(z[1:(cps3[i+1]-1)],)[2])
+ }
+ if(i>0&&i<length(cps3)){
+ MeanPM3<-c(MeanPM3,mean(zpm[cps3[i]:(cps3[i+1]-1)]))
+ MeanT3<-c(MeanT3,mean(zt[cps3[i]:(cps3[i+1]-1)]))
+ VarPM3<-c(VarPM3,var(zpm[cps3[i]:(cps3[i+1]-1)]))
+ VarT3<-c(VarT3,var(zt[cps3[i]:(cps3[i+1]-1)]))
+ Corr3<-c(Corr3,cor(z[cps3[i]:(cps3[i+1]-1)],)[2])
+ }
+ if(i==length(cps3)){
+ MeanPM3<-c(MeanPM3,mean(zpm[cps3[i]:720]))
+ MeanT3<-c(MeanT3,mean(zt[cps3[i]:720]))
+ VarPM3<-c(VarPM3,var(zpm[cps3[i]:720]))
+ VarT3<-c(VarT3,var(zt[cps3[i]:720]))
+ Corr3<-c(Corr3,cor(z[cps3[i]:720,])[2])
+ }
+ }
> MeanPM3
> MeanT3
> VarPM3
> VarT3
> Corr3
> pmz3<-c()
> pmf3<-c()
> tz3<-c()
> tf3<-c()
> for(i in 0:(length(cps3)-1)){
+ if(i==0){
+ x1<-zpm[1:(cps3[i+1]-1)]
+ y1<-zpm[cps3[i+1]:(cps3[i+2]-1)]
+ x2<-zt[1:(cps3[i+1]-1)]
+ y2<-zt[cps3[i+1]:(cps3[i+2]-1)]
+ z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+ f1<-var.test(x1,y1)
+ z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+ f2<-var.test(x2,y2)
+ pmz3<-c(pmz3,round(z1$p.value,4))
+ pmf3<-c(pmf3,round(f1$p.value,4))
+ tz3<-c(tz3,round(z2$p.value,4))
+ tf3<-c(tf3,round(f2$p.value,4))
+ }

```



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY


```

+ if(i>0&&i<(length(cps3)-1)){
+   x1<-zpm[cps3[i]:(cps3[i+1]-1)]
+   y1<-zpm[cps3[i+1]:(cps3[i+2]-1)]
+   x2<-zt[cps3[i]:(cps3[i+1]-1)]
+   y2<-zt[cps3[i+1]:(cps3[i+2]-1)]
+   z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+   f1<-var.test(x1,y1)
+   z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+   f2<-var.test(x2,y2)
+   pmz3<-c(pmz3,round(z1$p.value,4))
+   pmf3<-c(pmf3,round(f1$p.value,4))
+   tz3<-c(tz3,round(z2$p.value,4))
+   tf3<-c(tf3,round(f2$p.value,4))
+ }
+ if(i==length(cps3)-1){
+   x1<-zpm[cps3[i]:(cps3[i+1]-1)]
+   y1<-zpm[cps3[i+1]:720]
+   x2<-zt[cps3[i]:(cps3[i+1]-1)]
+   y2<-zt[cps3[i+1]:720]
+   z1<-z.test(x1,y1,sigma.x = var(x1),sigma.y = var(y1))
+   f1<-var.test(x1,y1)
+   z2<-z.test(x2,y2,sigma.x = var(x2),sigma.y = var(y2))
+   f2<-var.test(x2,y2)
+   pmz3<-c(pmz3,round(z1$p.value,4))
+   pmf3<-c(pmf3,round(f1$p.value,4))
+   tz3<-c(tz3,round(z2$p.value,4))
+   tf3<-c(tf3,round(f2$p.value,4))
+ }
+ }
> pmz3
> pmf3
> tz3
> tf3

##### Split Data #####
> data<-read.csv('dust.csv',sep=',',header=T)
> pm<-data$PM
> temp<-data$Temp
> z<-matrix(c(pm,temp),ncol=2)
> colnames(z)<-c("PM2.5","Temperature")
> z1<-matrix(c(pm[1:240],temp[1:240]),ncol=2)

```

```

> z2<-matrix(c(pm[241:480],temp[241:480]),ncol=2)
> z3<-matrix(c(pm[481:720],temp[481:720]),ncol=2)
> output11<-e.divisive(z1,R=499,alpha=1)
> output12<-e.divisive(z2,R=499,alpha=1)
> output13<-e.divisive(z3,R=499,alpha=1)
> output21<-e.cp3o(z1,K=300)
> output22<-e.cp3o(z2,K=300)
> output23<-e.cp3o(z3,K=300)
> output31<-ks.cp3o(z1,K=300)
> output32<-ks.cp3o(z2,K=300)
> output33<-ks.cp3o(z3,K=300)
> output4<-c()
> for(i in 1:3){
+   a<-240*(i-1)
+   cp1<-cpt.meanvar(z[,1][(1+a):(240+a)],method="PELT")
+   cp2<-cpt.meanvar(z[,2][(1+a):(240+a)],method="PELT")
+   cps<-sort(union(cpts(cp1),cpts(cp2)))
+   if(i==2||i==3){
+     cps<-sort(union(cpts(cp1),cpts(cp2))+(240*(i-1)))
+   }
+   output4<-c(output4,cps)
+ }
> output1<-c(output11$estimates[-c(1,length(output11$estimates))],output12$estimates[-
c(1,length(output12$estimates))+240,output13$estimates[-c(1,length(output13$estimates))+480])
> output2<-c(output21$estimates,output22$estimates+240,output23$estimates+480)
> output3<-c(output31$estimates,output32$estimates+240,output33$estimates+480)
> output1
> output2
> output3
> output4

```

ประวัติผู้เขียน

ชื่อ-สกุล	นาย ปภาวิน เจริญชัยปิยกุล
วัน เดือน ปี เกิด	9 พฤษภาคม 2539
สถานที่เกิด	กรุงเทพมหานคร
วุฒิการศึกษา	วิทยาศาสตรบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย
ที่อยู่ปัจจุบัน	40/7 ซอยอินทามระ 23 ถนนสุทธิสารวินิจฉัย แขวงสามเสนใน เขตพญาไท กรุงเทพมหานคร 10400



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY