การวิเคราะห์ตัวแบบค่าสุดขีดของปริมาณความเข้มข้น $PM_{2.5}$ ในเขตธุรกิจของกรุงเทพมหานคร

Analytical the extreme value model of the $PM_{2.5}$ concentration data

in the business area of Bangkok

นางสาวเอมิกา พูนวัฒนานุกูล

นางสาวมนิศา แย้มศรี

นางสาวหทัยชนก ลักษณา

โครงงานพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต

ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

โครงงานพิเศษ	การวิเคราะห์ตัวแบบค่าสุดขีดของปริมาณความเข้มข้น $PM_{2.5}$ ในเขตธุรกิจของกรุงเทพมหานคร Analytical the extreme value model of the $PM_{2.5}$ concentration data in the business area of Bangkok
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.คณิตา เพ็ชรัตน์
ผู้จัดทำโครงงานพิเศษ	1. นางสาว เอมิกา พูนวัฒนานุกูล รหัสนักศึกษา 5904052616343 2. นางสาว มนิศา แย้มศรี รหัสนักศึกษา 5904052636212 3. นางสาว หทัยชนก ลักษณา รหัสนักศึกษา 5904052636310
ภาควิชา	สถิติประยุกต์
คณะ	คณะวิทยาศาสตร์ประยุกต์
ปีการศึกษา	2562

คณะกรรมการตรวจสอบโครงงานพิเศษ ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ อนุมัติให้นับโครงงานพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต

คณะกรรมการตรวจสอบโครงงานพิเศษ

รองศาสตราจารย์ ดร.เสาวณิต สุขภารังษี กรรมการ

รองศาสตราจารย์ ดร.ยุพาภรณ์ อารีพงษ์ กรรมการ

ผู้ช่วยศาสตราจารย์ ดร.คณิตา เพ็ชรัตน์ อาจารย์ที่ปรึกษา

หัวข้อโครงงานพิเศษ	การวิเคราะห์ตัวแบบค่าสุดขีดของปริมาณความเข้มข้น $PM_{2.5}$ ในเขตธุรกิจของกรุงเทพมหานคร
	Analytical the extreme value model of the $PM_{2.5}$ concentration data in the business area of Bangkok
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.คณิตา เพ็ชรัตน์
ผู้จัดทำโครงงานพิเศษ	1. นางสาว เอมิกา พูนวัฒนานุกูล รหัสนักศึกษา 5904052616343
	2. นางสาว มนิศา แย้มศรี รหัสนักศึกษา 5904052636212
	3. นางสาว หทัยชนก ลักษณา รหัสนักศึกษา 5904052636310
ภาควิชา	สถิติประยุกต์
คณะ	คณะวิทยาศาสตร์ประยุกต์
ปีการศึกษา	2562

บทคัดย่อ

กรุงเทพมหานครเป็นเมืองหลวงของประเทศและเป็นมหานครที่เป็นศูนย์กลางความเจริญในหลายๆ ด้าน ก่อให้เกิดความหนาแน่นของประชากร มีการสร้างตึกสูง มีความคับคั่งของจราจร ส่งผลต่อมลพิษทางสภาพอากาศ และการเพิ่มสูงขึ้นของปริมาณความเข้มข้นของ $PM_{2.5}$ ในกรุงเทพมหานคร ณ ปัจจุบัน โดยในงานวิจัยนี้ ข้อมูลที่ใช้ในการศึกษาคือข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ใน 3 เขตธุรกิจของกรุงเทพมหานครซึ่งเป็นเขตที่ติดอันดับที่มีปริมาณความเข้มข้นของ $PM_{2.5}$ อยู่ในระดับเกินเกณฑ์มาตรฐานที่มีผลกระทบต่อสุขภาพของประชาชนที่ต้องทำงานหรือที่อาศัยอยู่บริเวณนั้น ได้แก่ เขตวังทองหลาง - สถานีสถานีตำรวจนครบาลโชคชัย (เดือนมิถุนายน ปี 2560 - เดือนกันยายน ปี 2562) และเขตดินแดง - สถานีการเคหะดินแดง (เดือนพฤษภาคม ปี 2554 - เดือนกันยายน ปี 2562) และเขตปทุมวัน - สถานีโรงพยาบาลจุฬาลงกรณ์ (เดือนพฤษภาคม ปี 2560 - เดือนกันยายน ปี 2562) นำมาใช้คาดการณ์ปริมาณความเข้มข้นของ $PM_{2.5}$ โดยเลือกใช้ตัวแบบที่มีความเหมาะสมของแต่ละสถานี เพื่อเป็นแนวทางในการรับมือและแก้ไขปัญหามลพิษทางอากาศใน 3 เขตของกรุงเทพมหานครในอนาคต ในงานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์เพื่อหาตัวแบบที่เหมาะสมจากทฤษฎีค่าสุดขีด (Extreme Value Theory: EVT) กับข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงและรายสัปดาห์แต่ละเขตภายใต้การแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution : GEV) และการแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution : GPD) โดยแบบจำลองที่ได้มานั้นสามารถคาดการณ์ระดับการเกิดซ้ำของค่าสุดขีดของปริมาณ

ความเข้มข้นของ $PM_{2.5}$ ในอีก 2 ปี 5 ปี 10 ปี 50 ปี และ 100 ปี ข้างหน้าของแต่ละสถานีได้ซึ่งในงานวิจัยนี้ใช้เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike's information criterion: AIC) และเกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criterion: BIC) เป็นเกณฑ์ในการเปรียบเทียบความเหมาะสมของตัวแบบ

กิตติกรรมประกาศ

โครงงานพิเศษเรื่อง แบบจำลองค่าสุดขีดของปริมาณความเข้มข้นของ $PM_{2.5}$ โดยได้รับความอนุเคราะห์ด้านข้อมูลในการศึกษาจากท่านผู้อำนวยการกรมควบคุมมลพิษ ขอขอบพระคุณ ผศ.ดร.ปิยภัทร บุษบาบดินทร์ ผศ.ดร.คณิตา เพ็ชรัตน์ อาจารย์ที่ปรึกษางานวิจัยฉบับนี้ ที่ได้เสียสละเวลาและให้คำแนะนำเรื่อง การวิเคราะห์ค่าสุดขีดด้วยโปรแกรม R รศ.ดร.เสาวณิต สุขภารังษี รศ.ดร.ยุพาภรณ์ อารีพงษ์ ที่เป็นกรรมการการสอบโครงงานพิเศษนี้ และการสนับสนุนจากภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ เพื่อให้งานวิจัยฉบับนี้สมบูรณ์ คณะผู้วิจัย ขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้ คณะผู้วิจัยหวังเป็นอย่างยิ่งว่าผลของการศึกษานี้จะเป็นประโยชน์ไม่มากก็น้อย สำหรับหน่วยงานที่เกี่ยวข้องตลอดจนผู้ที่สนใจ หากมีข้อผิดพลาดประการใด คณะผู้วิจัยขออภัยเป็นอย่างสูง

สุดท้ายนี้ คณะผู้วิจัยขอกราบขอบพระคุณบิดา มารดา และทุกคนในครอบครัว ที่คอยให้การสนับสนุนส่งเสริม เป็นกำลังใจตลอดปีการศึกษา และขอบคุณเพื่อนๆ พี่ๆ น้องๆ ในภาควิชาสถิติประยุกต์ รวมทั้งบุคคลที่มิได้กล่าวถึงมา ณ ที่นี่ด้วย ที่คอยเป็นกำลังใจ เสนอแนะติชมในการจัดทำงานวิจัยฉบับนี้มาโดยตลอดเป็นอย่างดี จนกระทั่งงานวิจัยฉบับนี้ได้สำเร็จลุล่วงไปได้ด้วยดี

นางสาว เอมิกา พูนวัฒนานุกูล
นางสาว มนิศา แย้มศรี
นางสาว หทัยชนก ลักษณา

สารบัญ

สารบัญ (ต่อ)

สารบัญ (ต่อ)

หน้า

สารบัญ (ต่อ)

หน้า

สารบัญภาพ

สารบัญภาพ (ต่อ)

ภาพที่ **หน้า**

สารบัญตาราง

ตารางที่ **หน้า**

สารบัญตาราง (ต่อ)

ตารางที่ **หน้า**

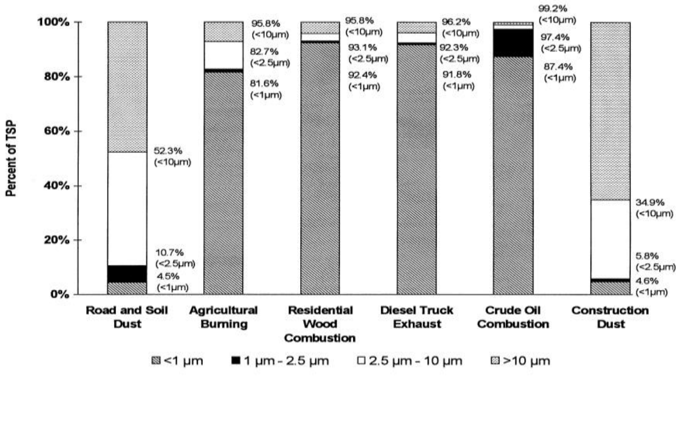
บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญ

กรุงเทพมหานครมีการเจริญเติบโตของเมืองอย่างต่อเนื่อง จากการเติบโตจากภาคเกษตรกรรมมาเป็นภาคอุตสาหกรรมทำให้กรุงเทพมหานครซึ่งเป็นศูนย์กลางของแหล่งธุรกิจและความเจริญมีจำนวนประชากรเพิ่มขึ้นอย่างรวดเร็ว มีการก่อสร้างอาคารและที่อยู่อาศัยมากขึ้น มีการใช้รถยนต์และรถจักรยานยนต์เพิ่มมากขึ้น การขยายตัวอย่างต่อเนื่องนั้นก่อให้เกิดการสะสมของปัญหาต่างๆ เช่น ปริมาณของขยะมูลฝอย การจราจรติดขัดเข้าขั้นวิกฤตและนับวันยิ่งทวีความรุนแรงมากขึ้น รวมถึงมลพิษทางอากาศซึ่งเป็นปัญหาที่ส่งผลกระทบต่อคุณภาพชีวิตของประชากรในกรุงเทพมหานครเป็นอย่างมาก

จากการสำรวจข้อมูลจากแอปพลิเคชั่น Air Visual ระบุว่า สภาพอากาศโดยทั่วไปของกรุงเทพมหานครตามมาตรฐาน US AQI อยู่ที่ 136 (PM2.5 อยู่ที่ 57.8 ไมโครกรัม/ลูกบาศก์เมตร) ติดอันดับที่ 12 เมืองที่มีคุณภาพอากาศแย่ที่สุดในโลก จากการติดตามตรวจสอบคุณภาพอากาศในพื้นที่กรุงเทพมหานครและปริมณฑล โดยกรมควบคุมมลพิษร่วมกับกรุงเทพมหานคร ทั้ง 46 สถานี พบว่า ปริมาณฝุ่นละอองตรวจวัดค่าได้ระหว่าง 37-67 ไมโครกรัมต่อลูกบาศก์เมตร (ค่ามาตรฐานไม่เกิน 50 ไมโครกรัมต่อลูกบาศก์เมตร) คุณภาพอากาศอยู่ในระดับดี 1 สถานีระดับปานกลาง 25 สถานี และระดับที่เริ่มส่งผลกระทบต่อสุขภาพ 20 สถานี ในกรุงเทพมหานคร ได้แก่ เขตบางขุนเทียน เขตดินแดง เขตสัมพันธวงศ์ เขตวังทองหลาง เขตปทุมวัน เขตบางคอแหลม เขตยานนาวา เขตคลองสาน เขตบางกอกน้อย เขตภาษีเจริญ เขตบางเขน เขตบางพลัด เขตสาทร เขตบางซื่อ เขตหลักสี่ และเขตบึงกุ่ม และพื้นที่ ตำบลปากน้ำ อำเภอเมือง, ตำบลทรงคนอง อำเภอพระประแดง และ ตำบลมหาชัย อำเภอเมือง จังหวัดสมุทรปราการ ดังนั้นข้อมูลคุณภาพอากาศจึงมีความสำคัญอย่างยิ่ง ไม่เพียงกระตุ้นประชาชนให้รับมือกับสถานการณ์ปัจจุบันและปกป้องสุขภาพของประชาชน แต่ยังสร้างความตื่นรู้ของสาธารณะ และขับเคลื่อนปฏิบัติการเพื่อต่อสู้กับมลพิษทางอากาศในระยะยาวของภาครัฐอีกด้วย



รูปภาพ 1-1 แหล่งกำเนิดของการกระจายของฝุ่นละออง (กรมควบคุมมลพิษ)

ปัญหามลภาวะจาก $PM_{2.5}$ จึงเป็นสิ่งที่ควรจัดการควบคู่ไปกับแผนพัฒนายุทธศาสตร์ชาติเพื่อพัฒนาคุณภาพอากาศไปสู่การพัฒนาที่ยั่งยืน โดยฝุ่นละอองจากแหล่งกำเนิดต่างๆนั้นจะมีขนาดที่แตกต่างกัน (รูปภาพ 1-1) ฝุ่นละอองจากถนนและดินฟุ้งปลิวรวมถึงฝุ่นจากการก่อสร้างจะมีขนาดใหญ่กว่า 2.5 ไมครอน มากกว่าร้อยละ 90 หรือน้อยกว่าร้อยละ 10 เป็นฝุ่นขนาดเล็กกว่า 2.5 ไมครอน หรือฝุ่น $PM_{2.5}$ ส่วนฝุ่นจากการเผาในที่โล่ง เช่น การเผาของเสียเกษตรกรรมซึ่งเป็นการเผาไหม้ที่ไม่สมบูรณ์อาจมีฝุ่นขนาดเล็กกว่า 2.5 ไมครอนประมาณร้อยละ 80 ในขณะที่การเผาไหม้น้ำมันดิบ ไอเสียรถยนต์ดีเซล และการเผาไม้ในเตาเผาครัวเรือนจะมีฝุ่นขนาดเล็กกว่า 2.5 ไมครอนมากกว่าร้อยละ 90 ดังนั้น หากเราพูดถึงฝุ่นไอเสียรถยนต์ย่อมหมายถึงฝุ่นขนาดเล็กกว่า 2.5 ไมครอนแต่การฉีดน้ำรดถนนหรือฉีดละอองน้ำในอากาศอาจช่วยลดฝุ่นขนาดใหญ่ในอากาศแต่มีประสิทธิภาพในการลดฝุ่นขนาดเล็กกว่า 2.5 ไมครอนน้อยมาก (กรมควบคุมมลพิษ, 2018)

ในการศึกษาครั้งนี้จะมุ่งศึกษาเกี่ยวกับข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ใน 3 เขตธุรกิจของกรุงเทพมหานครที่กำลังเผชิญกับปัญหามลพิษทางอากาศที่เกิดขึ้นมาอย่างต่อเนื่อง ได้แก่ เขตดินแดงที่ตั้งอยู่ทางฝั่งตะวันออกของแม่น้ำเจ้าพระยาหรือฝั่งพระนคร มีสภาพทั่วไปเป็นแหล่งการค้า การบริการ และแหล่งที่อยู่อาศัยหนาแน่นเป็นจำนวนมาก เขตปทุมวัน เป็นเขตหนึ่งที่ตั้งอยู่บริเวณใจกลางที่สุดของกรุงเทพมหานครและที่มีการคมนาคมหลากหลายช่องทาง เป็นเขตศูนย์กลางธุรกิจ การค้า การบริการ การพยาบาล วัฒนธรรม การศึกษาและการทูต เขตวังทองหลาง มีสภาพโดยทั่วไปเป็นแหล่งที่อยู่อาศัยหนาแน่นปานกลาง โดยมีย่านการค้าหนาแน่นทางด้านตะวันตกเฉียงเหนือของพื้นที่

(1) (2) (3)

รูปภาพ 1- 2 แสดงแผนที่ (1) เขตดินแดง (2) เขตปทุมวันและ (3) เขตวังทองหลาง

โดยมลพิษทางอากาศใน 3 เขต ส่วนใหญ่มีสาเหตุหลักมาจากการจราจรที่หนาแน่นและการคมนาคมจากจำนวนรถยนต์และรถขนส่งสาธารณะในการเดินทางที่เพิ่มมากขึ้นการสร้างตึกและอาคารรวมถึง

ที่อยู่อาศัยอย่างแออัด เป็นสาเหตุสำคัญที่เกิดขึ้นตามการเจริญเติบโตและการพัฒนาของเมือง ดังนั้น $PM_{2.5}$ จึงเป็นมลพิษทางอากาศที่ได้รับความสนใจเพื่อควบคุมคุณภาพอากาศให้อยู่ในระดับที่พอดีไม่ส่งผลกระทบต่อสุขภาพของประชาชน อีกทั้งยังต้องพร้อมรับมือกับการเปลี่ยนแปลงสภาพอากาศที่ทำให้อากาศนิ่งและมีสภาพอากาศปิดทำให้มีการสะสมของฝุ่นละอองในอากาศเพิ่มขึ้น ประกอบกับมีหมอกทำให้ไม่มีการไหลเวียนของอากาศ ฝุ่นละอองจึงลอยขึ้นสู่บรรยากาศ (กรมควบคุมมลพิษ, 2018) ทำให้ปริมาณความเข้มข้นของ $PM_{2.5}$ อยู่ในระดับที่สูงเกินกว่าค่ามาตรฐาน ซึ่งโดยทั่วไปผู้วิจัยส่วนใหญ่จะใช้เทคนิคการพยากรณ์ด้วยอนุกรมเวลาโดยการใช้วิธีแยกส่วนประกอบ (Decomposition Method) วิธีพยากรณ์ของวินเตอร์ (Winter's Method) และเทคนิควิเคราะห์อนุกรมเวลาของบอกซ์-เจนกินส์ (Box-Jenkins) โดยเครื่องมือทางสถิติที่จะเข้ามามีบทบาทเกี่ยวข้องในเรื่องนี้ คือการวิเคราะห์ค่าสุดขีด (Extreme Value Analysis) (ปิยภัทรและอรุณ, 2015) ซึ่งเป็นเครื่องมือจากการประยุกต์ใช้ทฤษฎีค่าสุดขีดที่กล่าวถึงคุณสมบัติของเหตุการณ์ที่มีตัวแปรสุ่มจัดอยู่ในลักษณะที่เรียกว่า "ค่าสุดขีด" ที่อาจจะเป็นค่าสูงสุดหรือต่ำสุดก็ได้ พร้อมทั้งศึกษารูปแบบการแจกแจงความน่าจะเป็นของตัวแปรสุ่มเหล่านี้ การวิเคราะห์ข้อมูลเมื่อข้อมูลมีค่าสุดขีด (Extreme Value) นั้น ส่วนใหญ่จะตัดข้อมูลส่วนนี้ทิ้งไปไม่นำมาพิจารณา เนื่องจากมีความซับซ้อนและยุ่งยากในการวิเคราะห์ แต่ในความเป็นจริงหากต้องการทราบถึงความน่าจะเป็นในการเกิดขึ้นของเหตุการณ์ที่มีค่าสูงสุดหรือต่ำสุด ซึ่งอยู่ในส่วนของปลายหางซึ่งมีค่าน้อยมาก จะต้องนำข้อมูลค่าสุดขีดมาวิเคราะห์ด้วยเพื่อช่วยคาดการณ์ความน่าจะเป็นของข้อมูลที่ศึกษาภายในระยะเวลาที่กำหนดและสามารถประมาณความเสียหายสุดท้ายที่จะเกิดขึ้น ตลอดจนการสร้างแบบจำลองเหตุการณ์ที่เกิดขึ้นไม่บ่อยหรือเหตุการณ์บางอย่างที่เกิดขึ้นได้ยาก (Embrechts et al, 1997) นั่นคือ ช่วยในการตัดสินใจหากมีแนวโน้มที่จะเกิดขึ้นเหตุการณ์ที่สุดขีดนั้นขึ้นอีก

การประยุกต์ใช้การแจกแจงพาเรโตวางนัยทั่วไป (GPD) เริ่มต้นโดย Pickand (1975) และมีงานวิจัยวิเคราะห์ค่าสุดขีดของข้อมูลมลพิษในอากาศและเปรียบเทียบระหว่างสองเขตมหานครในอเมริกาใต้ภายใต้การแจกแจงพาเรโตวางนัยทั่วไป (GPD) และการแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) ในข้อมูลปริมาณความเข้มข้นของ CO, NO, NO_2, PM_{10} และ $PM_{2.5}$ (Martins et al, 2017) จากที่กล่าวมาข้างต้น คณะผู้วิจัยได้เล็งเห็นถึงความสำคัญของปริมาณความเข้มข้นของ $PM_{2.5}$ ใน 3 เขตธุรกิจของกรุงเทพมหานคร ซึ่งเป็นเขตที่ติดอันดับที่มีปริมาณความเข้มข้นของ $PM_{2.5}$ อยู่ในระดับสูงที่มีผลกระทบต่อสุขภาพของประชาชนที่ต้องทำงานหรือที่อาศัยอยู่บริเวณนั้น ได้แก่ เขตดินแดง เขตปทุมวัน และเขตวังทองหลาง โดยใช้ทฤษฎีค่าสุดขีด (Extreme Value Theory) กับข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงและรายสัปดาห์ของแต่ละเขต โดยการเปรียบเทียบด้วยการแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution: GEV) และ

การแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution: GPD) โดยแบบจำลองที่ได้มานั้นสามารถคาดการณ์ระดับการเกิดซ้ำของปริมาณความเข้มข้นของ $PM_{2.5}$ ในอีก 2 ปี 5 ปี 10 ปี 50 ปี และ 100 ปี ข้างหน้า

1.2 วัตถุประสงค์

1.2.1 เพื่อวิเคราะห์หาตัวแบบค่าสุดขีดที่เหมาะสมของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ โดยใช้การแจกแจงค่าสุดขีดนัยทั่วไป (GEV) ใน 3 เขตธุรกิจของกรุงเทพมหานคร

1.2.2 เพื่อวิเคราะห์หาตัวแบบค่าสุดขีดที่เหมาะสมของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ โดยใช้การแจกแจงพาเรโตนัยทั่วไป (GPD) ใน 3 เขตธุรกิจของกรุงเทพมหานคร

1.2.3 เพื่อเปรียบเทียบความเหมาะสมของตัวแบบระหว่างการแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) และการแจกแจงพาเรโตวางนัยทั่วไป (GPD) ใน 3 เขตธุรกิจของกรุงเทพมหานครโดยใช้เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike’s information criterion: AIC) และเกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criterion: BIC) เป็นเกณฑ์ในการเปรียบเทียบความเหมาะสมของตัวแบบ

1.2.4 เพื่อหาระดับการเกิดซ้ำของปริมาณความเข้มข้นของ $PM_{2.5}$ ใน 3 เขตธุรกิจของกรุงเทพมหานคร

1.3 ขอบเขตงาน

1.3.1 โครงงานพิเศษนี้มุ่งศึกษาตัวแบบที่เหมาะสมจากทฤษฎีค่าสุดขีดกับข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ใน 3 เขตธุรกิจของกรุงเทพมหานคร

1.3.2 ข้อมูลที่ใช้ในการศึกษา คือ ข้อมูลทุติยภูมิ (Secondary Data) ข้อมูลรายชั่วโมงและรายสัปดาห์ของปริมาณความเข้มข้นของ $PM_{2.5}$ จากสถานีสถานีตำรวจนครบาลโชคชัย เขตวังทองหลาง (เดือนมิถุนายน ปี 2560 ถึง เดือนกันยายน ปี 2562) สถานีการเคหะดินแดง เขตดินแดง (เดือนพฤษภาคม ปี 2554 ถึง เดือนกันยายน ปี 2562) และสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน (เดือนพฤษภาคม ปี 2560 ถึง เดือนกันยายน ปี 2562)

1.4 นิยาม

ค่าสุดขีด คือ ตัวแปรสุ่มซึ่งอาจจะเป็นค่าสูงสุดหรือต่ำสุดก็ได้และเป็นค่านอกเกณฑ์ที่มีค่าที่น้อยกว่า ควอไทล์ที่ 1 หรือมากกว่าควอไทล์ที่ 3 เกิน 3 เท่าของค่าพิสัยระหว่างควอไทล์ (IQR)

มลพิษทางอากาศ คือ ภาวะอากาศที่มีส่วนประกอบของอากาศเปลี่ยนแปลงไปมีปริมาณของฝุ่นละออง ก๊าซ กลิ่น หมอกควัน ไอ ไอน้ำ เขม่าและกัมมันตภาพรังสีอยู่ในบรรยากาศปริมาณที่สูงกว่า

ระดับปกติเป็นเวลานานพอที่จะทำให้เกิดอันตรายแก่มนุษย์ สัตว์ พืช หรือทรัพย์สินต่างๆ มลพิษอาจเกิดขึ้นเองตามธรรมชาติ กรณีที่เกิดจากการกระทำของมนุษย์ ได้แก่ มลพิษจากท่อไอเสียของพาหนะ โรงงานอุตสาหกรรม ขบวนการผลิต เป็นต้น จากปัญหามลพิษต่างๆมีการปล่อยสารมลพิษที่กระทบต่อสิ่งแวดล้อมและมนุษย์ได้แก่ ก๊าซไนโตรเจนไดออกไซต์ (NO_x) ก๊าซไฮโดรคาร์บอน (HC) ก๊าซคาร์บอนมอนอกไซต์ (CO) และฝุ่นละออง (Particulate Matters: PM) ซึ่งส่งผลกระทบต่อสุขภาพของประชากรเป็นบริเวณกว้าง

$PM_{2.5}$ คือ ฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน เทียบอย่างง่ายว่ามีขนาดประมาณ 1 ใน 25 ของเส้นผ่านศูนย์กลางของเส้นผมมนุษย์ ขนจมูกของมนุษย์ไม่สามารถกรองได้ จึงสามารถแพร่กระจายเข้าสู่ทางเดินหายใจ กระแสเลือด และแทรกซึมสู่กระบวนการทำงานของอวัยวะต่างๆ ในร่างกายเพิ่มความเสี่ยงเป็นโรคเรื้อรังและมะเร็ง ตามคำเตือนขององค์การอนามัยโลก

1.5 ผลที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางในการวางแผนการจัดการแก้ปัญหาปริมาณความเข้มข้นของ $PM_{2.5}$ เพื่อควบคุมคุณภาพอากาศและรับมือการเกิดมลพิษทางอากาศที่มีค่าเกณฑ์เกินมาตราฐานที่จะส่งผลกระทบต่อสุขภาพของประชาชนที่ดำเนินชีวิตอยู่ใน 3 เขตธุรกิจของกรุงเทพมหานคร

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาวิเคราะห์ทางสถิติเพื่อหาตัวแบบที่เหมาะสมจากทฤษฎีค่าสุดขีด (Extreme Value Theory: EVT) กับข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงและรายสัปดาห์ของแต่ละเขต ผู้วิจัยได้ทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องประกอบด้วยหัวเรื่องดังต่อไปนี้

2.1 เอกสารที่เกี่ยวข้อง
- 2.1.1 ความคงที่ของข้อมูล (Stationary)
- 2.1.2 ความมีแนวโน้มของข้อมูล (Trend)
- 2.1.3 ทฤษฎีค่าสุดขีด (Extreme Value Theory)
- 2.1.4 การแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized extreme value: GEV)
- 2.1.5 การแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution: GPD)
- 2.1.6 การตรวจสอบความเหมาะสมของการแจกแจง
- 2.1.7 ระดับการเกิดซ้ำ (Return Level)

2.2 งานวิจัยที่เกี่ยวข้อง

มีรายละเอียดดังต่อไปนี้

2.1 เอกสารที่เกี่ยวข้อง

2.1.1 ความคงที่ของข้อมูล (Stationary)

คุณสมบัติของอนุกรมเวลามี 2 ประเภท คือ อนุกรมเวลาที่อยู่ภายใต้กระบวนการคงที่ (Stationary process) และอนุกรมเวลาที่อยู่ภายใต้กระบวนการไม่คงที่ (Non-Stationary process) ซึ่งความคงที่ (Stationarity) หมายความว่า ค่าเฉลี่ย (Mean) ค่าความแปรปรวน (Variance) และค่าความแปรปรวนร่วม (Covariance) จะไม่ผันแปรไปตามเวลาที่เปลี่ยนไป (Enders, 2004) โดยจะตรวจสอบความคงที่ของข้อมูลอนุกรมเวลได้จากกราฟเส้น แผนภาพ Correlogram และทดสอบ Unit root ของข้อมูลนั้นว่ามีความคงที่หรือไม่ ด้วยวิธี Augmented Dickey-Fuller (ADF) test เป็นต้น

โดยงานวิจัยนี้เลือกวิธีที่ใช้ทดสอบความคงที่ของข้อมูล คือ Augmented Dickey-Fuller (ADF) test ที่เสนอโดย Dickey and Fuller 1979 และ 1981 เนื่องจากเป็นวิธีที่ได้รับการยอมรับและเป็นที่นิยมอย่างแพร่หลายในการศึกษาความคงที่ของข้อมูลอนุกรมเวลา และหากผลการทดสอบ

ได้แสดงให้เห็นว่าข้อมูลมีความไม่คงที่ นั่นคือ ชุดของข้อมูลเหล่านี้มีการเคลื่อนไหวไปตามแนวโน้มที่เพิ่มขึ้นตามกาลเวลา (Time Trend) และความแปรปรวนวิ่งห่างออกจากเดิมไปเรื่อยๆตามแนวโน้มของระยะเวลาที่เพิ่มขึ้น

มีสมการดังต่อไปนี้

การทดสอบ Unit Root โดยวิธีการ Augmented Dickey Fuller เป็นสมการที่ปรับมาจากวิธีการ Dickey Fuller ซึ่งสมการของ Dickey Fuller มี 3 สมการ ที่ต้องทดสอบ (At Level) คือ

$\triangle Y_t = \delta Y_{t-1} + u_t$ (Random Walk Process) (1)

$\triangle Y_t = \beta_1 + \delta Y_{t-1} + u_t$ (Random Walk with Drift) (2)

$\triangle Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + u_t$ (Random Walk with Drift around a Stochastic Trend) (3)

เมื่อมีการปรับสมการ โดยการเพิ่มขบวนการถดถอยในตัวเอง (Autoregressive Processes) เข้าไปในสมการ 1, 2 และ 3 จะได้สมการดังนี้

$$\triangle Y_t = \delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \triangle Y_{t-1} + u_t \quad (4)$$

$$\triangle Y_t = \delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \triangle Y_{t-1} + u_t \quad (5)$$

$$\triangle Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \triangle Y_{t-1} + u_t \quad (6)$$

โดยมีสมมติฐาน

$$H_0 : \delta = 0$$
$$H_a : \delta \neq 0$$

ถ้ายอมรับสมมติฐานหลัก (H_0) แสดงว่า Y_t มีลักษณะไม่นิ่ง (Nonstationary) หรือเรียกว่ามี Unit Root เมื่อค่าสถิติ t-Statistic ของสัมประสิทธิ์ในรูปสัมบูรณ์มีค่าน้อยกว่าค่าวิกฤต Mackinnon Critical Value

ถ้าปฏิเสธสมมติฐานหลัก (H_0) หรือยอมรับสมมติฐานรอง (H_a) แสดงว่า Y_t มีลักษณะนิ่ง (Stationary) หรือเรียกว่าไม่มี Unit Root เมื่อค่าสถิติ t-Statistic ของสัมประสิทธิ์ในรูปสัมบูรณ์มีค่ามากกว่าค่าวิกฤต Mackinnon Critical Value

2.1.2 ความมีแนวโน้มของข้อมูล (Trend)

ความมีแนวโน้ม หมายถึง ความเปลี่ยนแปลงของข้อมูลที่มีลักษณะราบเรียบแนวโน้มอาจมีลักษณะเป็นเส้นตรงหรือเส้นโค้งในทางเพิ่มขึ้นหรือลดลง ค่าแนวโน้มของข้อมูลเป็นการเคลื่อนไหวในช่วงระยะเวลาที่ค่อนข้างนานพอสมควรโดยวิธีที่ใช้ทดสอบคือวิธี Mann-Kendall (Mann, 1945) และ (Kendall, 1975) ซึ่งเป็นการทดสอบทางสถิติแบบไม่มีพารามิเตอร์ ที่นิยมอย่างมาก สำหรับการที่จะนำมาใช้วิเคราะห์หาแนวโน้มของอนุกรมเวลาของข้อมูลทางด้านอุทกวิทยา (Yue และ Wang, 2004) ในการจัดกลุ่มแนวโน้มที่มีนัยสำคัญของแนวโน้มที่เพิ่มขึ้นและแนวโน้มที่ลดลง และสำหรับอนุกรมเวลา $X = (\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_n)$ สมมติฐานหลัก (Null Hypothesis : H_0) คือ ตัวอย่างของ n เป็นตัวแปรสุ่มที่ไม่ต้องการเงื่อนไขความเป็นอิสระและการแจกแจงเดียวกัน สมมติฐานรอง (Alternative Hypothesis : H_1) สำหรับการทดสอบแบบสองทิศทางจะมีการแจกแจงของ x_k และ x_j ซึ่งจะมีค่าไม่เหมือนกันกับทุกค่าของ k โดย $j \le n$ ด้วย $k \ne j$ โดยสถิติทดสอบ S สามารถคำนวณได้โดยสมการนี้ (Kahya & Kalayci, 2004)

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sign(x_j - x_k)$$

$$\text{เมื่อ } sign(x_j - x_k) = \begin{cases} 1; (x_j - x_k) > 0 \\ 0; (x_j - x_k) = 0 \\ -1; (x_j - x_k) < 0 \end{cases}$$

และค่าเฉลี่ยและค่าความแปรปรวนของสถิติทดสอบ S สามารถหาได้จาก

$$E(\mathrm{S}) = 0$$

$$\mathrm{Var(S)} = \frac{\left[n(n-1)(2n+5) - \sum_{j=1}^{m} t_j (t_j - 1)(2t_j + 5) \right]}{18}$$

การแจกแจงความโน้มเอียงของ S และแนวโน้มที่มีนัยสำคัญสามารถทดสอบได้ โดยการเปรียบเทียบค่าตัวแปรมาตรฐาน z จากสมการ

$$z = \begin{cases} \dfrac{S-1}{\sqrt{Var(S)}} & ; S > 0 \\ 0 & ; S = 0 \\ \dfrac{S+1}{\sqrt{Var(S)}} & ; S < 0 \end{cases}$$

ซึ่งค่าบวกของ z ในสมการ เป็นการแสดงถึงแนวโน้มที่เพิ่มขึ้น ในขณะที่ค่าลบของ z แสดงถึงแนวโน้มที่ลดลง

2.1.3 ทฤษฎีค่าสุดขีด (Extreme Value Theory)

ทฤษฎีค่าสุดขีด (Extreme Value Theorem) เป็นทฤษฎีที่กล่าวถึงคุณสมบัติของเหตุการณ์ที่มีตัวแปรสุ่ม ซึ่งจัดอยู่ในลักษณะที่เรียกว่า “ค่าสุดขีด” อาจจะเป็นค่าสูงสุดหรือต่ำสุดก็ได้ พร้อมทั้งศึกษารูปแบบการแจกแจงความน่าจะเป็นของตัวแปรสุ่มเหล่านี้ ลักษณะของการวิเคราะห์แบบจำลองค่าสุดขีดด้วยทฤษฎีค่าสุดขีด สามารถแบ่งลักษณะการแจกแจงของค่าสุดขีดได้เป็น 2 ประเภทตามลักษณะของการเลือกข้อมูลค่าสุดขีดที่นำมาวิเคราะห์ ได้แก่ การแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution: GEV) และการแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution: GPD) ซึ่งการสร้างแบบจำลองด้วย GEV เหมาะสำหรับวิเคราะห์ค่าสุดขีดในช่วงคาบเวลาที่สนใจ เช่น รายปี รายเดือน รายไตรมาส หรือรายสัปดาห์ เป็นต้น ซึ่งค่าสังเกตที่รวบรวมไว้ควรจะมีจำนวนมากกว่า 30 ปีขึ้นไป โดยจะเลือกข้อมูลที่สูงสุดในแต่ละช่วงคาบเวลา (Block time) ที่ผู้วิเคราะห์สนใจ แต่ถ้าต้องการวิเคราะห์การแจกแจงของปลายหางของข้อมูลเหล่านี้เมื่อข้อมูลมีจำนวนมาก หรือ ข้อมูลเก็บรวบรวมเป็นรายวัน การสร้างแบบจำลองด้วย GPD จะมีความเหมาะสมกว่า GEV เนื่อง GPD จะอธิบายลักษณะข้อมูลที่มีการแจกแจงแบบมีหางหนา (Heavy-tailed Distribution) ได้ดีกว่า และจำนวนค่าสุดขีดที่ นำมาวิเคราะห์ด้วย GPD จะมีจำนวนมากกว่าข้อมูลที่นำมาวิเคราะห์ด้วย GEV ซึ่งสามารถลดความไม่แน่นอน ที่เกิดขึ้นจากการสุ่มตัวอย่างได้ การสร้างแบบจำลอง ด้วย GPD มีขั้นตอนสำคัญ คือ การกำหนดค่าเกณฑ์ (Threshold) ที่เหมาะสมกับข้อมูลที่นำมาวิเคราะห์และการพิจารณาความไม่เป็นอิสระของข้อมูลค่าสุดขีดซึ่งสามารถแก้ไขได้โดยการจัดกลุ่มค่าสุดขีด (Declustering) ที่มีค่าเกินกว่าค่าเกณฑ์

โดยส่วนใหญ่ในการวิเคราะห์ข้อมูลถ้าพบข้อมูลที่เป็นค่าสุดขีด (Extreme Value) จะไม่นำข้อมูล ส่วนนั้นมาใช้ในการวิเคราะห์อาจจะทำให้การวิเคราะห์เกิดความแปรปรวนและคลาดเคลื่อนได้ แต่ถ้าต้องการทราบเกี่ยวกับความน่าจะเป็นของเหตุการณ์ที่มีค่าสูงสุดหรือต่ำสุดที่อยู่ในส่วนของปลายหางจะมีค่าน้อยมาก เช่น ปริมาณน้ำที่ไหลลงอ่างสูงสุด-ต่ำสุดในแต่ละสัปดาห์ ปริมาณน้ำฝนสูงสุด-ต่ำสุดในแต่ละเดือน และความเร็วลมสูงสุดในรอบเดือนหรืออุณหภูมิสูงสุด-ต่ำสุดในแต่ละวัน เป็นต้น ทฤษฎีค่าสุดขีด (EVT) คือ การศึกษาเกี่ยวกับความน่าจะเป็นของเหตุการณ์สุดขีดที่คาดว่าจะเกิดขึ้น ในอนาคตหรือการสร้างแบบจำลองที่ดีที่สุดเพื่อใช้ในการศึกษาข้อมูล โดยส่วนใหญ่ในการศึกษาวิเคราะห์ข้อมูลที่มีอยู่จากการเก็บรวบรวมข้อมูลจะให้ผลของข้อมูลเป็นค่าต่างๆ ดังนั้น การหาแบบจำลองที่ดีที่สุดสาหรับข้อมูลที่ต้องการศึกษาเป็นขั้นตอนที่มีบทบาทหน้าที่ต่อการวางแผนและตัดสินใจผลลัพธ์ที่ได้จะขึ้นอยู่กับเหตุการณ์ที่จะเกิดขึ้นในอนาคต (ปิยภัทร, 2017)

ซึ่งจากทฤษฎีรูปแบบค่าสุดขีด (Extremal Types Theorem) ถ้ามีค่าคงที่ลำดับ $a_n > 0$ และ b_n เมื่อ $n \to \infty$ จะพบว่า

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq G(\mathrm{x})\right\}$$

สำหรับบางการแจกแจงที่เป็นการแจกแจงแบบไม่แปรสภาพ (Non-degenerate distribution) แล้วจะทำให้มีคุณสมบัติเป็นการแจกแจงอย่างใดอย่างหนึ่งของการแจกแจงต่อไปนี้

$$G(\mathrm{x}) = \exp\left\{-\exp\left(-x\right)\right\} : -\infty < x < \infty$$
$$G(\mathrm{x}) = \begin{cases} 0 : x \leq 0 \\ \exp\left\{-x^{-a}\right\} : x > 0, a > 0 \end{cases}$$
$$G(\mathrm{x}) = \begin{cases} \exp\left\{-(\mathrm{x}-)^{a}\right\} : x <, a > 0 \\ 1 : x \geq 0 \end{cases}$$

การแจกแจงทั้งสามชนิดของทฤษฎีค่าสุดขีดเป็นที่รู้จักกันในชื่อของกัมเบล (Gumbel) ฟรีเชท (Frechet) และ ไวล์บูล (Weibull) ตามลำดับ โดยเรียกการแจกแจงทั้งสามตัวนี้ว่า “การแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution: GEV)” สำหรับการแจกแจงกัมเบลและฟรีเซทเป็นการแจกแจงที่ G ไม่มีขอบเขตล่าง แต่สำหรับการแจกแจงไวล์บูลเป็นการแจกแจงที่มีขอบเขตบน และเป็นที่น่าสังเกตว่าไม่สามารถระบุลิมิตของ M_n ด้วยทฤษฎีค่าสุดขีด ถึงแม้ว่าจะระบุชนิดของการแจกแจงแล้วก็ตาม ดังนี้ $\frac{M_n - b_n}{a_n} \xrightarrow{D} G$ ซึ่งถ้าหาการแจกแจงได้ จะสามารถระบุชนิดของการแจกแจงว่าเป็นการแจกแจงค่าสุดขีดชนิดใด

ดังนั้นทฤษฎีค่าสุดขีด (EVT) เป็นทฤษฎีที่อธิบายถึงคุณสมบัติของเหตุการณ์ที่มีตัวแปรสุ่มซึ่งจัดอยู่ในลักษณะที่เรียกว่า “ค่าสุดขีด” อาจจะเป็นค่าสูงสุดหรือต่ำสุดก็ได้ พร้อมทั้งศึกษารูปแบบการแจกแจงความน่าจะเป็นของตัวแปรสุ่มเหล่านี้

ลักษณะของการวิเคราะห์แบบจำลองค่าสุดขีดด้วยทฤษฎีค่าสุดขีด (EVT) การแจกแจงของค่าสุดขีดยังสามารถแบ่งลักษณะได้ 2 ประเภทตามลักษณะของการเลือกข้อมูลค่าสุดขีดที่นำมาวิเคราะห์ คือ

1) การแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution: GEV)

2) การแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution: GPD)

โดยการสร้างแบบจาลองด้วยการแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) เหมาะสำหรับวิเคราะห์ ค่าสุดขีดในช่วงคาบเวลาที่สนใจ เช่น รายปี รายเดือน รายไตรมาส หรือรายสัปดาห์ เป็นต้น และค่าสังเกตที่รวบรวมไว้ควรจะมีค่าสังเกตมากกว่า 30 ปี ขึ้นไป โดยจะเลือกข้อมูลที่สูงสุดในแต่ละช่วงคาบเวลา (Block time) ที่ผู้วิจัยสนใจ

2.1.3.1 การแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized extreme value: GEV)

การแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) เหมาะสำหรับวิเคราะห์ค่าสุดขีดในช่วงคาบเวลาที่สนใจ เช่น รายปี รายเดือน รายไตรมาส หรือรายสัปดาห์ เป็นต้น ซึ่งค่าสังเกตที่รวบรวมได้ ควรจะมีจำนวนมากกว่า 30 ปี ขึ้นไป โดยจะเลือกข้อมูลที่สูงสุดในแต่ละช่วงคาบเวลาที่ผู้วิเคราะห์สนใจ (Susanti, Adnan, Yendra, และ Muhaijir, 2018)

ให้ X_i เมื่อ $i = 1, 2, ..n$ เป็นตัวแปรสุ่มที่อิสระต่อกันและมีฟังก์ชันการแจกแจงสะสม $F(\mathrm{x};\theta)$ แบบเดียวกัน กำหนดให้ค่าสูงสุดของตัวแปรสุ่ม คือ $X_{(\mathrm{n})} = Max(\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_n)$ ซึ่งประยุกต์ใช้ในรูปแบบของการแจกแจงค่าสุดขีดวางนัยทั่วไป ที่มีพารามิเตอร์ 3 ตัวแปร ได้แก่ (Coles, 2001)

1) μ แทนพารามิเตอร์ระบุตำแหน่ง (Location Parameter)
2) σ แทนพารามิเตอร์ระบุขนาด (Scale Parameter)
3) ξ แทนพารามิเตอร์ระบุรูปร่าง (Shape Parameter)

ถ้าค่าคงที่ลำดับ $a_n > 0$ และ b_n ดังนั้น

$$\Pr\left\{\frac{M_n - b_n}{a_n} \le Z\right\} \xrightarrow{D} G(\mathrm{z})$$

เมื่อ $n \to \infty$ สำหรับฟังก์ชัน G เป็นการแจกแจงแบบไม่แปรสภาพ (Non-degenerate Distribution)

จะสามารถกล่าวได้ว่าเป็นสมาชิกของการแจกแจง GEV ดังสมการ G

$$F(\mathrm{x};\mu,\sigma,\xi) = \exp\left\{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right\}$$

บนเงื่อนไข $\left\{z : 1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0\right\}$ เมื่อ $-\infty < \mu < \infty, \sigma > 0$ และ $-\infty < \xi < \infty$ ที่มีฟังก์ชันการแจกแจงความน่าจะเป็น (Probability Distribution Function : PDF) ดังนี้

$$f(\mathrm{x}) = \frac{1}{\sigma}\left\{1+\xi\left(\frac{x-\mu}{\sigma}\right)\right\}^{(-1/\xi)-1} \exp\left\{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right\}$$

สำหรับ $1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0$

จากสมการที่พบว่า

เมื่อ $\xi = 0$ เรียกการแจกแจงค่าสุดขีดวางนัยทั่วไปว่า “การแจกแจงกัมเบล”

เมื่อ $\xi > 0$ เรียกการแจกแจงค่าสุดขีดวางนัยทั่วไปว่า “การแจกแจงฟรีเชท”

และ เมื่อ $\xi < 0$ เรียกการแจกแจงค่าสุดขีดวางนัยทั่วไปว่า “การแจกแจงไวล์บูล”

อาจจะเรียกทฤษฎีนี้ว่า “ทฤษฎีสามแบบ” (Three Types Theorem) ตามลักษณะของลิมิตการแจกแจงฟังก์ชัน $F(\mathrm{x};\theta)$ (Galambos, 1978)

แบบที่ 1 รูปแบบกัมเบล (Gumbel Type)

$F(\mathrm{x})\exp(-\exp(-x)$ เมื่อ $-\infty < x < \infty$

แบบที่ 2 รูปแบบฟรีเชท (Fréchet Type)

$F(\mathrm{x})\exp(-x-a)$ เมื่อ $-\infty < x < \infty$ และ $a > 0$ พบว่าถ้า $F(\mathrm{x}) = 0$ เมื่อ $\mathrm{x} < 0$

แบบที่ 3 รูปแบบไวล์บูล (Weibull Type)

$F(\mathrm{x})\exp(-x-a)$ เมื่อ $-\infty < x < \infty$ และ $a > 0$ พบว่า $F(\mathrm{x}) = 1$ เมื่อ $\mathrm{x} > 0$

2.1.3.2 การแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution: GPD)

เป็นการแจกแจงที่นิยมใช้เมื่อลักษณะข้อมูลมีการแจกแจงแบบหางหนา (Heavy-tailed Distribution) การแจกแจงพาเรโตวางนัยทั่วไปโดยมี 2 พารามิเตอร์ คือ σ แทนพารามิเตอร์ระบุขนาด (Scale Parameter) และ ξ แทนพารามิเตอร์ระบุรูปร่าง (Shape Parameter) โดยมีฟังก์ชันการแจกแจงความน่าจะเป็น (PDF) เป็นดังนี้

$$f(\mathrm{x}) = 1+\left[\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}$$

และมีฟังก์ชันการแจกแจงสะสม (CDF) เป็นดังนี้

$$\mathrm{F(x)} = 1-\left(1+\frac{\xi x}{\sigma}\right)^{-1/\xi}$$

เมื่อ $\sigma > 0$ และ $-\infty < \xi < \infty$

โดยพิจารณาจากค่าประมาณของพารามิเตอร์ระบุรูปร่าง (ξ) จะบอกลักษณะข้อแตกต่างของการแจกแจงแต่ละแบบ ดังนี้

ถ้า $\xi < 0$ การแจกแจงของค่าเกินเกณฑ์จะมีขอบเขตบนหรือการแจกแจงพาเรโต II

ถ้า $\xi > 0$ การแจกแจงของค่าเกินเกณฑ์จะไม่มีลิมิตบนหรือการแจกแจงพาเรโตทั่วไป

ถ้า $\xi = 0$ การแจกแจงจะมีขอบบนและล่าง และเมื่อ $\xi \to 0$ จะได้

$$F(\mathrm{x}) = 1 - \exp\left(\frac{-x}{\hat{\sigma}}\right), x > 0$$

คือการแจกแจงแบบเลขชี้กำลังที่มีค่าคาดหวัง $= \frac{1}{\hat{\sigma}}$

2.1.3.3 วิธีเกินเกณฑ์ (Threshold methods)

เป็นวิธีที่นิยมใช้เพื่อเลือกค่าสุดขีดที่จะนำมาวิเคราะห์ซึ่งวิธีนี้จะเลือกค่าสังเกตที่มีค่าสูงกว่าค่าที่กำหนด หรือเรียกว่า ค่าเกณฑ์ (Threshold) การเลือกค่าเกณฑ์ที่เหมาะสม คือ เลือกค่าเกณฑ์ที่มีความคงที่ (Threshold stability choice) หมายถึง การสร้างแบบจำลอง GPD ที่เป็นไปได้สำหรับค่าที่มากกว่าค่าเกณฑ์ (u_0) บางค่า แล้วจะมีค่าที่มากกว่าค่าเกณฑ์ทั้งหมด หรือ $u > u_0$ ถ้ากำหนดให้ $\sigma\mu_0$ เป็นพารามิเตอร์ระบุขนาดของค่าที่มากกว่าค่า μ_0 ที่มีการแจกแจงแบบการแจกแจงพาเรโตวางนัยทั่วไป ดังนั้น ค่าคาดหวังของค่าที่สูงกว่าค่าเกณฑ์จะอยู่ภายใต้เงื่อนไขต่อไปนี้

$\mathrm{E}[\mathrm{X} - u \mid \mathrm{X} > u] = \dfrac{\sigma u + \xi u}{1 - \xi}$ นั่นคือทุกๆ $u > u_0$ เมื่อ $\mathrm{E}[\mathrm{X} - u \mid \mathrm{X} > u]$ ซึ่งเป็นฟังก์ชันเส้นตรงของ u นอกจากนั้น $\mathrm{E}[\mathrm{X} - u \mid \mathrm{X} > u]$ เป็นค่าเฉลี่ยของค่าที่มีมากกว่าค่า u และสามารถประมาณค่าเฉลี่ยตัวอย่างของค่าที่สูงกว่าค่า μ ด้วยการประมาณจากกราฟความคลาดเคลื่อนเฉลี่ย (Mean residual life plot: MRL) ซึ่งสร้างจากค่าเกณฑ์ที่สูงและมีความเหมาะสมที่จะนำมาสร้างแบบจำลองการแจกแจงพาเรโตวางนัยทั่วไป

2.1.4 การประมาณภาวะความน่าจะเป็นสูงสุด (Maximum Likelihood Estimation: MLE)

ในการประมาณค่าพารามิเตอร์ที่ต้องการสำหรับการแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution : GEV) และการแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution: GPD) ทั่วไปนั้นจะใช้การประมาณค่าภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation: MLE) เนื่องจากให้ค่าความแกร่งมากที่สุด ซึ่งจากฟังก์ชันการแจกแจงความน่าจะเป็นของ GEV และ GPD สามารถเขียนฟังก์ชันล็อคไลค์ลิฮูดได้ดังนี้

$$g(x,\mu,\sigma,\xi)=\frac{1}{\sigma}\left\{1+\xi\left(\frac{x-\mu}{\sigma}\right)\right\}^{(-1/\xi)-1}\exp\left\{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right\}$$

2.1.5 การตรวจสอบความเหมาะสมของแบบจำลอง

2.1.5.1 Akaike Information Criterion (AIC)

เป็นดัชนีที่ใช้ในทฤษฎีข้อมูลซึ่งมีว่าหากจะให้แบบจำลองสามารถอธิบายข้อมูลได้ดีขึ้น อาจจำเป็นที่จะต้องเพิ่มตัว parameter ขึ้นมา ซึ่งทำให้แบบจำลองมีความซับซ้อนมากขึ้นในการเลือกอาไคเคะ (Akaike, 1973) ได้เสนอเกณฑ์การคัดเลือกตัวแบบคือเกณฑ์ AIC ซึ่งสร้างจากการประมาณความแปรปรวนของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback-Leibler Information) ระหว่างตัวแบบที่แท้จริง (True Model) กับตัวแบบที่เหมาะสมที่มีคุณสมบัติไม่เอนเอียงเมื่อขนาดตัวอย่างใหญ่ การคัดเลือกตัวแบบที่ให้ AIC มีค่าต่ำที่สุดซึ่งจะเป็นตัวแบบที่ดีที่สุด และเกณฑ์ AIC สามารถคัดเลือกตัวแบบได้ดีเมื่อตัวอย่างมีขนาดใหญ่เนื่องจากเกิดความผิดพลาดในการคัดเลือกตัวแบบสูงเมื่อตัวอย่างมีขนาดเล็ก (McQuarrie, Shumway และ Tsai, 1997) ซึ่งเกณฑ์ AIC สามารถนิยามได้โดย

$$AIC=-2In(\mathrm{L})+2\mathrm{m}$$

โดยที่ L เป็นฟังก์ชันภาวะความน่าจะเป็นของการแจกแจง (likelihood function)

m เป็นจำนวนพารามิเตอร์ของการแจกแจง

2.1.5.2 Bayesian Information Criteria (BIC)

เกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criteria: BIC) ถูกนำเสนอโดยซาวา (Sawa, 1978) โดยพัฒนามาจากเกณฑ์ AIC ทั้งนี้เกณฑ์ BIC จะเลือกตัวแบบที่ให้ค่า BIC ต่ำสุดเป็นตัวแบบที่ถูกต้อง โดยมีสูตร ดังนี้ (Muller et al.,2013)

$$BIC = -2In(\mathrm{L}) + \mathrm{m}\ln(\mathrm{n})$$

โดยที่ L เป็นฟังก์ชันภาวะความน่าจะเป็นของการแจกแจง (likelihood function)

m เป็นจำนวนพารามิเตอร์ของการแจกแจง

n เป็นจำนวนข้อมูล

2.1.5.3 The Komolgorov-simrnov test (KS-Test)

การทดสอบแบบนี้เป็นวิธีทดสอบภาวะรูปดี (Goodness of fit test) อีกวิธีหนึ่ง เช่นเดียวกันกับการทดสอบไคสแควร์ แต่เป็นวิธีที่มีประสิทธิภาพมากกว่าและใช้ได้กับทุกข้อมูลทุกกรณี ถึงแม้ว่าความถี่บางกลุ่มจะเป็น 0 ก็ตาม การคำนวณค่าสถิติใน Kolmogorov-Smirnov One-Sample test จะใช้ความถี่สะสมแทนความถี่ปกติ (ทั้งความถี่ที่สังเกตได้ และความถี่คาดหวัง) จุดมุ่งหมายของการทดสอบเช่นเดียวกับไคสแควร์ คือ ต้องการทดสอบว่าการแจกแจงของข้อมูลที่สังเกตได้คล้ายคลึง (เท่ากัน) กับการแจกแจงที่คาดหวังหรือไม่

2.1.5.4 Quantile-Quantile plots (Q-Q plots)

การสร้างกราฟ Q-Q คือ การเปรียบเทียบค่าควอนไทล์ (Quantile) ของแบบจำลองกับค่าข้อมูลจริง กล่าวคือ การที่พล็อตจุดระหว่าง

$$\left\{\left(\widehat{F}^{-1}\frac{i}{n+1}, x_{(i)}\right) : i = 1, 2, ..., n\right\}$$

ถ้าแบบจำลองที่ได้มีการแจกแจงเหมือนกับข้อมูลจริง กราฟที่ได้จะมีลักษณะเป็นเส้นตรง ซึ่งการสร้างกราฟนี้เป็นการพล็อตจุดควอนไทล์ของค่าคาดหวังของการแจกแจงที่ได้จากการสร้างแบบจำลองกับจุดควอนไทล์ของข้อมูลจริงที่ต้องการทดสอบว่าการแจกแจงมีความเหมาะสมหรือไม่ ซึ่งการประมาณฟังก์ชันการแจกแจง $\widehat{F}$ ถ้าข้อมูลที่ศึกษามีการแจกแจงที่เหมาะสม Q-Q plots จะอยู่ใกล้เส้นทแยงมุมซึ่งกราฟจะเป็นเส้นตรง

2.1.6 ระดับการเกิดซ้ำ (Return Level)

การประมาณค่าระดับการเกิดซ้ำ คือการประมาณสิ่งที่จะเกิดขึ้นหนึ่งครั้งในระยะเวลาทุกๆ หน่วยเวลาที่สนใจ เรียกว่า “ระดับการเกิดซ้ำ” (Return Levels) หลังจากตรวจสอบความแม่นยำของแบบจำลองที่สร้างขึ้นจะประมาณระดับการเกิดซ้ำ r ปี แทนด้วย Z_r สาหรับช่วงเวลาใดๆ ที่ต้องการ ถ้าให้ฟังก์ชันระดับการเกิดซ้ำของ GEV คือ $1-1/\mathrm{r}$ เมื่อ $\mathrm{x}=\hat{\mathrm{Z}}_r$ แทนค่าประมาณระดับการเกิดซ้ำ สูงสุดในรอบ r ปี การสร้างแบบจำลอง GEV คือ

$$1-G(\hat{\mathrm{Z}}_{100}:\hat{\mu},\hat{\sigma},\hat{\xi}) \quad \text{ซึ่งคือ} \quad 1-\exp\left\{-\left(1+\hat{\xi}\left(\frac{\hat{\mathrm{Z}}_{100}-\hat{\mu}}{\hat{\sigma}}\right)\right)^{-1/\xi}\right\}=0.01$$

และสามารถประมาณค่าระดับการเกิดซ้ำ r ปี โดยแทนด้วย $\hat{\mathrm{Z}}_r$ ตามสมการต่อไปนี้

$$\hat{\mathrm{Z}}_r=\hat{\mu}+\frac{\hat{\sigma}}{\hat{\xi}}\left[(-\log(0.09))^{-\xi}-1\right]$$

2.2 งานวิจัยที่เกี่ยวข้อง

ทฤษฎีค่าสุดขีดถูกนำมาประยุกต์ใช้ในการวิจัยด้านต่าง ๆ ดังนี้

พิมพ์พรรณ (2017) งานวิจัยนี้ได้ศึกษาข้อมูลทั่วไปเกี่ยวกับค่าแนวโน้มน้ำฝนสะสมสูงสุดและต่ำสุดด้วยทฤษฎีความเป็นไปได้ของแม่น้ำท่าจีนในจังหวัดสุพรรณบุรีด้วยทฤษฎีค่าสุดขีดด้วยการแจกแจงค่าสุดขีดวางนัยทั่วไปและการแจกแจงพาเรโตวางนัยทั่วไป

นายชินวัฒน์ เมืองแก้ว และคณะ (2018) เพื่อวิเคราะห์หาแบบจำลองที่เหมาะสมจากทฤษฎีค่าสุดขีด (Extreme Value Theory) สำหรับข้อมูลปริมาณน้ำ ที่ไหลลงอ่างเก็บน้ำสูงสุดในแต่ละสัปดาห์ของเขื่อนขุนด่านปราการชล ตั้งแต่ปี พ.ศ. 2550-2561 ด้วยการแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution)

นางสาวนัชชา แก้วใจรักษา และคณะ (2017) ได้ศึกษาเกี่ยวกับการสร้างแบบจำลองที่เหมาะสมสำหรับข้อมูลปริมาณน้ำฝนของกรุงเทพมหานครด้วยการแจกแจงพาเรโตวางนัยทั่วไป

พัณณิ์ ภาริษา และคณะ (2013) ได้ทำการศึกษาหาตัวแบบที่เหมาะสมสำหรับปริมาณน้ำฝนสูงสุดของภาคเหนือตอนบน โดยใช้การแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution: GEV) พร้อมทั้งหาระดับการเกิดซ้ำ (Return levels) ของปริมาณน้ำฝนสูงสุดในรอบปีการเกิดซ้ำ (Return period)

วิชุดา และ ปิยภัทร (2017) จำลองอัตราการป่วยโรคไข้หวัดใหญ่ในภาคตะวันออกเฉียงเหนือของประเทศไทยโดยการวิเคราะห์แบบจำลองค่าสุดขีดแบ่งตามลักษณะการแจกแจง 2 ประเภท คือ GEV และ GPD

บทที่ 3

วิธีการดำเนินงาน

ในส่วนของวิธีการดำเนินงานวิจัยเรื่องแบบจำลองค่าสุดขีดของปริมาณความเข้มข้นของ $PM_{2.5}$ ใน 3 เขตของกรุงเทพมหานคร ได้แก่ เขตดินแดง เขตปทุมวัน และเขตวังทองหลาง จะเริ่มจากการวิเคราะห์ข้อมูลโดยพื้นฐานทางสถิติ แล้วจึงทำการทดสอบความคงที่(stationary) และแนวโน้มของข้อมูล (Trend) และหาตัวแบบจำลองโดยการประมาณค่าพารามิเตอร์ที่เหมาะสม จากนั้นทำการเลือกตัวแบบที่เหมาะสมโดยพิจารณาด้วยเกณฑ์ข้อสนเทศของอาไคเค (Akaike's Information Criterion : AIC) และเกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criteria : BIC) มีการทดสอบด้วยกราฟควอนไทล์ (Quantile-Quantile : Q-Q plots) จากการประมาณค่าพารามิเตอร์ด้วยวิธี Maximum Likelihood Estimation หรือ MLE พอได้ตัวแบบที่เหมาะสมแล้วจึงทำการประมาณค่าระดับการเกิดซ้ำเป็นขั้นตอนสุดท้าย

3.1 การเก็บรวบรวมข้อมูล

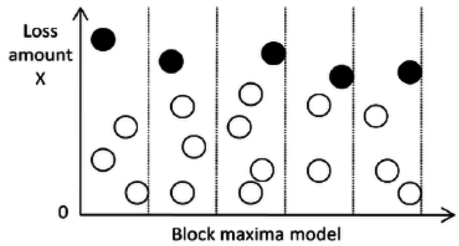
การศึกษานี้ใช้ข้อมูลทุติยภูมิ (Secondary Data) ซึ่งคือข้อมูลรายชั่วโมงและรายสัปดาห์ของปริมาณความเข้มข้นของ $PM_{2.5}$ใน 3 เขตธุรกิจของกรุงเทพมหานคร ได้แก่ สถานีสถานีตำรวจนครบาลโชคชัย - เขตวังทองหลาง (เดือนมกราคม ปี 2553 - เดือนกันยายน ปี 2562) สถานีการเคหะดินแดง - เขตดินแดง (เดือนมกราคม ปี 2553 - เดือนกันยายน ปี 2562) และสถานีโรงพยาบาลจุฬาลงกรณ์ - เขตปทุมวัน (เดือนมกราคม ปี 2553 - เดือนกันยายน ปี 2562)

3.2 การเลือกข้อมูล

เป็นการศึกษาความน่าจะเป็นในการเกิดขึ้นของเหตุการณ์ที่มีค่าสูงสุดหรือต่ำสุดซึ่งอยู่ในส่วนของปลายหางซึ่งมีวิธีการอยู่ 2 วิธี ได้แก่

3.2.1 Block maxima

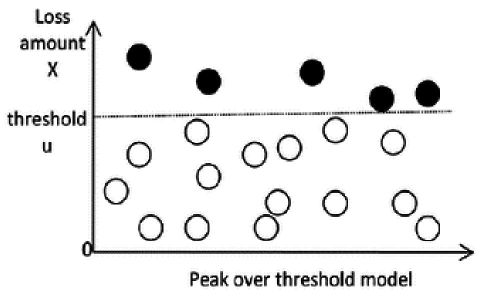
ใช้กับข้อมูลที่เป็นคาบเวลาที่สนใจ เช่น สัปดาห์ เดือน หรือปี เป็นต้น และพิจารณาค่าสูงสุดที่เกิดขึ้นในช่วงเวลานั้นๆ ซึ่งในงานวิจัยนี้จะใช้วิธี block maxima ในการเลือกข้อมูลเพื่อนำไปวิเคราะห์หาตัวแบบจำลองสำหรับวิธีการแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution: GEV) โดยหาค่าสูงสุดจากข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายวันของแต่ละเขตแล้วนำมาเก็บข้อมูลเป็นรายสัปดาห์



รูปภาพ 3-1 แสดงตัวแบบมีค่าสูงสุดในช่วงเวลา

3.2.2 Threshold exceedance

ใช้กับข้อมูลที่มีปริมาณมากหรือปลายหางหนา และอยู่ในคาบเวลาที่สนใจ เช่น รายชั่วโมง รายวัน เป็นต้น โดยจะพิจารณาร่วมกับวิธีเกินเกณฑ์ (Threshold methods) และใช้ข้อมูลที่มีค่าสูงเกินกว่าค่าเกณฑ์ (Threshold) ไปวิเคราะห์เพื่อหาตัวแบบ ซึ่งในงานวิจัยนี้จะใช้วิธี Threshold exceedance ในการเลือกข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงในแต่ละเขต เพื่อนำไปวิเคราะห์หาตัวแบบจำลองสำหรับการแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution : GPD) โดยกำหนดค่าเกณฑ์ (Threshold) ให้มีค่าไม่ต่ำกว่าเปอร์เซ็นไทล์ที่ 95% โดยจะพิจารณาค่าเปอร์เซ็นไทล์จากเงื่อนไขที่ว่า “จำนวนข้อมูลต้องอยู่ในช่วง 150 – 200 ค่า”



รูปภาพ 3-2 แสดงตัวแบบที่มีค่าเกินเกณฑ์

3.3 การวิเคราะห์ข้อมูล

3.3.1 วิเคราะห์ข้อมูลพื้นฐานทางสถิติ

งานวิจัยครั้งนี้ศึกษาคุณสมบัติของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงและรายสัปดาห์ ใน 3 เขตธุรกิจของกรุงเทพมหานคร ได้แก่ เขตดินแดง เขตปทุมวัน และเขตวังทองหลาง เพื่อนำมาวิเคราะห์ข้อมูลพื้นฐาน ได้แก่ ค่าต่ำสุด (Minimum) ค่าสูงสุด (Maximum) ค่าเฉลี่ย (Mean) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) และควอไทล์ (Quantile)

3.3.2 การทดสอบความคงที่ (Stationary) และแนวโน้ม (Trend) ของข้อมูล

จากข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงและรายสัปดาห์ของทั้ง 3 เขต ได้แก่ เขตดินแดง เขตปทุมวัน และเขตวังทองหลาง เป็นข้อมูลอนุกรมเวลา (Times Series Data) งานวิจัยนี้เลือกทดสอบความความคงที่ (Stationary) ของข้อมูลอนุกรมเวลาโดยใช้วิธีการทดสอบของ Augmented Dickey-Fuller (ADF) test และทดสอบแนวโน้มของข้อมูลอนุกรมเวลาใช้วิธีการทดสอบของ Mann-Kendall (MK) test ซึ่งวิธีการทดสอบนี้เหมาะสำหรับทดสอบแนวโน้มของข้อมูล ที่ไม่ได้มีการแจกแจงแบบปกติ โดยงานวิจัยนี้ใช้การวิเคราะห์ความคงที่และแนวโน้มอย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น 95% (P-value < 0.05)

3.3.3 การประมาณค่าพารามิเตอร์ที่เหมาะสม

กำหนดแบบจำลองที่เหมาะสมสำหรับข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของแต่ละเขตด้วยการแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution : GPD) และข้อมูลรายสัปดาห์ของแต่ละเขตด้วยการแจกแจงแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution : GEV) โดยจะกำหนดตัวแบบจำลองตามรูปแบบของข้อมูล ได้แก่
เมื่อข้อมูลมีความคงที่ (Stationary) และไม่มีแนวโน้ม (Trend) จะมีแบบจำลองได้ 1 รูปแบบ ได้แก่

แบบจำลองที่ 1 : μ, σ และ ξ เป็นค่าคงที่

และเมื่อข้อมูลไม่คงที่ (Non-Stationary) และมีแนวโน้ม (Trend) จะมีแบบจำลองของ ได้ 5 รูปแบบ ได้แก่

แบบจำลองที่ 1 : μ, σ และ ξ เป็นค่าคงที่

แบบจำลองที่ 2 : $\mu(t) = \beta_0 + \beta_1 t$ โดยที่ σ และ ξ เป็นค่าคงที่

แบบจำลองที่ 3 : $\mu(t) = \alpha_0 + \alpha_1 t$ โดยที่ μ และ ξ เป็นค่าคงที่

แบบจำลองที่ 4 : $\sigma(t) = \exp(\alpha_0 + \alpha_1 t)$ โดยที่ μ และ ξ เป็นค่าคงที่

แบบจำลองที่ 5 : $\mu(t) = \beta_0 + \beta_1 t$, $\sigma(t) = \exp(\alpha_0 + \alpha_1 t)$ โดยที่ ξ เป็นค่าคงที่

เมื่อ t มีหน่วยเป็นรายสัปดาห์ และ ξ เป็นค่าคงที่ของตัวแบบจำลองทั้งหมด โดยที่ ξ เป็นฟังก์ชันที่เปลี่ยนไปตามเวลา จึงทำให้การประมาณค่าพารามิเตอร์ไม่แม่นยำ (Coles, 2001) ดังนั้นในงานวิจัยนี้ประมาณค่าพารามิเตอร์สำหรับการแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution : GPD) และ การแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution : GEV) โดยพารามิเตอร์ μ, σ และ ξ จากทุกแบบจำลอง โดยใช้วิธีการประมาณค่าภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation : MLE)

3.3.4 การหาช่วงความเชื่อมั่นของพารามิเตอร์

จากการประมาณค่าพารามิเตอร์และความคลาดเคลื่อนมาตรฐานของการแจกแจง สามารถหาช่วงความเชื่อมั่นของพารามิเตอร์ด้วยวิธีวัลด์ (Wald Method) ซึ่งคำนวณจาก

$$\hat{\theta} \pm 1.96(s.e)$$

โดยที่ $\hat{\theta}$ คือ ค่าประมาณของพารามิเตอร์

$s.e$ คือ ค่าความคลาดเคลื่อนมาตราฐานของค่าประมาณพารามิเตอร์

ซึ่งในงานวิจัยนี้ ตรวจหาการแจกแจงที่เหมาะสมของตัวแบบจำลองด้วยการนำช่วงความเชื่อมั่นของพารามิเตอร์รูปร่างของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงในแต่ละเขตเทียบกับเงื่อนไขการแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution : GPD) และข้อมูลรายสัปดาห์ของแต่ละเขตเทียบกับเงื่อนไขการแจกแจงแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution : GEV)

3.3.5 การตรวจสอบความเหมาะสมของแบบจำลองค่าขีดสุด

ตรวจสอบแบบจำลองที่เหมาะสมกับข้อมูลโดยใช้เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike’s Information Criterion) และเกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criteria) โดยแบบจำลองที่เหมาะสมที่สุดนั้นจะมีค่า AIC และ BIC น้อยที่สุดและจะคัดเลือกตัวแบบได้ดีมากขึ้นตามขนาดของตัวอย่าง (Akaike, 1973) พร้อมทั้งใช้การทดสอบด้วยกราฟควอนไทล์ (Quantile-Quantile: Q-Q plots) ร่วมพิจารณาโดยการ Plot Quantile ของการแจกแจงหนึ่งกับ Quantile ของอีกการแจกแจงหนึ่ง ถ้าการแจกแจงทั้งสองเหมือนกันแล้วกราฟที่ได้จากการ Plot จะมีกราฟที่ใกล้เคียงกับเส้นตรงเป็นการ Plot Quantile ของค่าคาดหวังของการแจกแจงที่ได้กับ Quantile ของข้อมูลที่ต้องการทดสอบว่าการแจกแจงมีความเหมาะสมหรือไม่ในกรณีที่รูปกราฟไม่ fit กับเส้นตรงแสดงว่า ข้อมูลที่สนใจมีการแจกแจงไม่เหมาะสม ดังนั้น จึงอาจใช้การแปลงข้อมูลแล้ว ลอง Plot กราฟดูใหม่ว่าข้อมูลใหม่ที่เกิดจากการแปลงข้อมูลเดิมนั้นมีการแจกแจงที่เหมาะสมหรือไม่

3.3.6 การประมาณค่าระดับการเกิดซ้ำ (Return Level) หรือ ($\hat{Z}_r$)

ประมาณค่าระดับการเกิดซ้ำ (Return Level) หรือ ($\hat{Z}_r$) ของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงและรายสัปดาห์ของทั้ง 3 เขตของกรุงเทพมหานคร ได้แก่ เขตดินแดง เขตปทุมวัน และเขตวังทองหลาง ที่คาดว่าจะเกิดขึ้นโดยเฉลี่ยทุกๆช่วงเวลา t เมื่อ t คือรอบเวลาของการเกิดซ้ำ ซึ่งในงานวิจัยนี้กำหนดช่วงเวลาเป็น 2 ปี 5 ปี 10 ปี 50 ปี และ 100 ปี ข้างหน้า การศึกษาระดับการเกิดซ้ำนั้นมีความสำคัญในการคาดการณ์และการวางแผน เพื่อที่จะสามารถหาค่าประมาณพารามิเตอร์จากแบบจำลอง

3.4 โปรแกรมที่ใช้ในการวิเคราะห์และประมวลผลข้อมูล

ในงานวิจัยนี้ได้ทำการวิเคราะห์และประมวลเพื่อหาตัวแบบจำลองค่าสุดขีดที่เหมาะสมสำหรับข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ใน 3 เขตของกรุงเทพมหานครด้วยโปรแกรม R Studio Version 1.2.1335

บทที่ 4
ผลการดำเนินงาน

การดำเนินการศึกษาครั้งนี้มีวัตถุประสงค์เพื่อหาแบบจำลองที่เหมาะสมของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมง และรายสัปดาห์ของทั้ง 3 เขตของกรุงเทพมหานคร ได้แก่ เขตดินแดง เขตปทุมวัน และเขตวังทองหลาง ด้วยการแจกแจงพาเรโตวางนัยทั่วไป (GPD) และการแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution: GEV) โดยทำการวิเคราะห์ข้อมูลพื้นฐาน ทดสอบความคงที่ (Stationary) และการมีแนวโน้มของข้อมูล (Trend) กำหนดตัวแบบโดยการประมาณค่าพารามิเตอร์ที่เหมาะสม ตรวจสอบความเหมาะสมของตัวแบบหาระดับการเกิดซ้ำและการเปรียบเทียบตัวแบบที่มีความเหมาะสมจากทั้งสองการแจกแจงด้วยใช้เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike’s information criterion : AIC) และเกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criteria : BIC)

ขอบเขตเวลาของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ใน 3 เขตของกรุงเทพมหานคร รายชั่วโมงและรายสัปดาห์ ได้แก่ เขตดินแดง เขตปทุมวัน และเขตวังทองหลาง พบว่ามีบางสถานีที่มีข้อมูลสูญหายไปดังแสดงในตารางที่ 4-1 ดังนี้

ตารางที่ 4-1 แสดงขอบเขตเวลาของข้อมูลที่นำมาศึกษา

สถานี	ขอบเขตเวลาที่เอามาศึกษา	เหตุผล
สถานีสถานีตำรวจนครบาลโชคชัย เขตวังทองหลาง	7 มิถุนายน ปี 2560 ถึง 30 กันยายน ปี 2562	เครื่องมือขัดข้อง
สถานีการเคหะดินแดง เขตดินแดง	15 พฤษภาคม ปี 2554 ถึง 30 กันยายน ปี 2562	เครื่องมือขัดข้อง
สถานีโรงพยาบาลจุฬาลงกรณ์เขตปทุมวัน	1 มิถุนายน ปี 2560 ถึง 30 กันยายน ปี 2562	เครื่องมือขัดข้อง

4.1 ตัวแบบจำลองการแจกแจงพาเรโตวางนัยทั่วไป (GPD) จากปริมาณความเข้มข้นของ PM2.5 รายชั่วโมงของ 3 เขตธุรกิจในกรุงเทพมหานคร

จากผลการวิเคราะห์เพื่อหาแบบจำลองที่เหมาะสมของปริมาณความเข้มข้นของ $PM_{2.5}$ ใน 3 เขตของกรุงเทพมหานคร โดยมีข้อมูลพื้นฐานรายชั่วโมงผลสรุปดังตารางที่ 4-2 ต่อไปนี้

ตารางที่ 4-2 ข้อมูลรายชั่วโมงพื้นฐานของปริมาณความเข้มข้นของ $PM_{2.5}$ ใน 3 เขตธุรกิจในกรุงเทพมหานคร

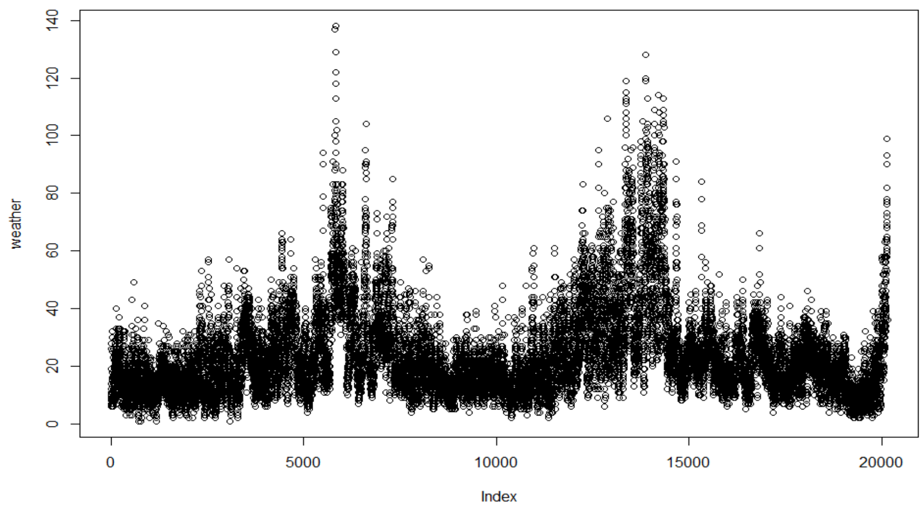
สถานี	จำนวนข้อมูล	ค่าสูงสุด	ค่าเฉลี่ย	S.D	ค่าต่ำสุด
สถานีสถานีตำรวจนครบาลโชคชัย เขตวังทองหลาง	20146	138	23.76	15.32	1
สถานีการเคหะดินแดง เขตดินแดง	47913	202.64	35.06	20.59	0.74
สถานีโรงพยาบาลจุฬาลงกรณ์เขตปทุมวัน	20105	130	26.54	15.04	3

(ไมโครกรัม/ลูกบาศก์เมตร (μg/m3))

4.1.1 สถานีสถานีตำรวจนครบาลโชคชัย เขตวังทองหลาง

4.1.1.1 การเตรียมข้อมูล

1. นำข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของสถานีสถานีตำรวจนครบาลโชคชัย เขตวังทองหลาง นำมาแสดงผลดังรูปภาพที่ 4-1



รูปภาพ 4-1 กราฟแสดงการกระจายของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของสถานีสถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลาง

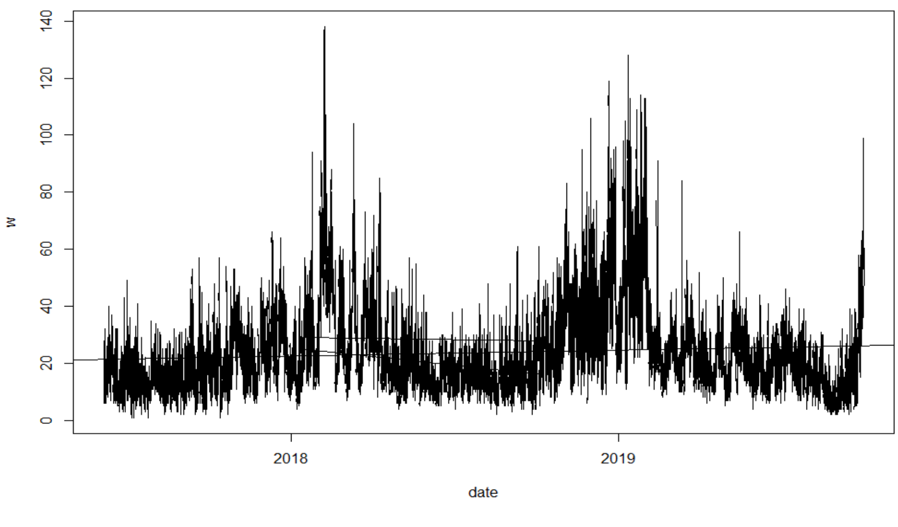
2. การทดสอบความคงที่ (Stationary) และแนวโน้ม (Trend) ของข้อมูล

ตารางที่ 4-3 การทดสอบความคงที่และแนวโน้มของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ของสถานีสถานีตำรวจนครบาลโชคชัย เขตวังทองหลาง

การทดสอบ	ค่าสถิติ	P-value
Augmented Dickey-Fuller (ADF) test	-8.9339	0.01
Mann-Kendall trend test	13.798	2.2e-16

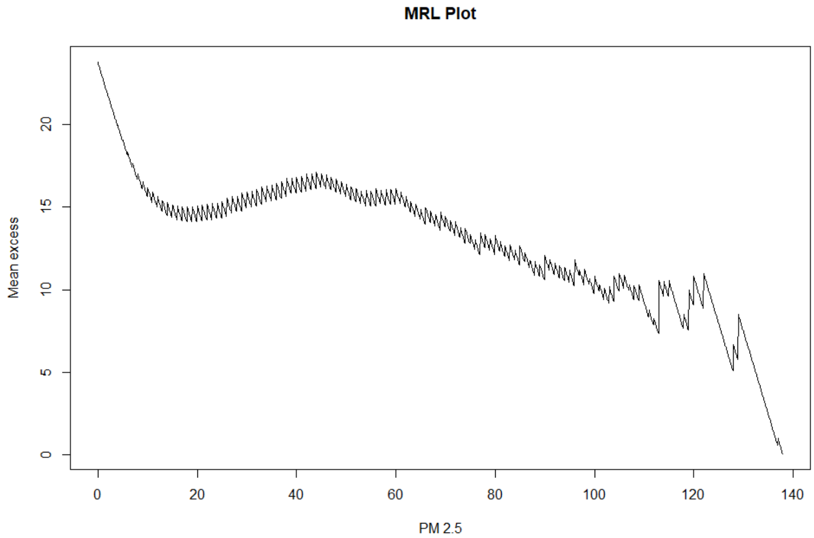
*(*ระดับนัยสำคัญ 0.05)*

จากตารางที่ 4-3 แสดงการทดสอบ จากวิธีการทดสอบของ Augmented Dickey-Fuller (ADF) test ที่ระดับนัยสำคัญ 0.05 เมื่อพิจารณาค่า P-value เท่ากับ 0.01 พบว่าข้อมูลชุดนี้อยู่ภายใต้กระบวนการคงที่ และพิจารณาจากวิธีทดสอบของ Mann-Kendall trend (MK) test ที่ระดับนัยสำคัญ 0.05 เมื่อพิจารณาค่า P-value เท่ากับ 2.2e-16 พบว่าข้อมูลชุดนี้ไม่มีแนวโน้มของข้อมูล แสดงดังรูปภาพที่ 4-2 ดังนี้

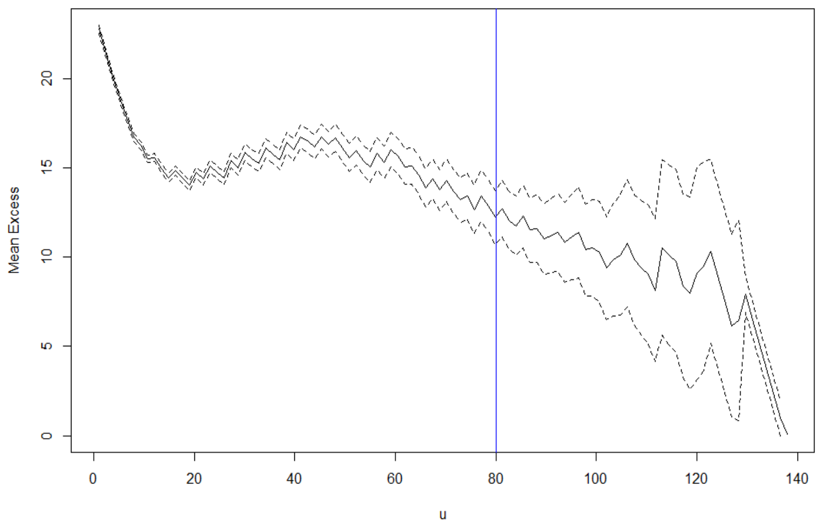


รูปภาพ 4-2 กราฟแสดง Time series ของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของสถานีสถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลาง

3. ทำการเลือกค่าเกินเกณฑ์ (μ_0) จากกราฟค่าความคลาดเคลื่อนเฉลี่ย (mean residual life plot : MRL) ดังรูปภาพที่ 4-3 โดยกำหนดค่าเกณฑ์ที่เปอร์เซ็นไทล์ 99% มีค่าเท่ากับ 80 เนื่องจากมีจำนวนข้อมูลที่ค่าเกินค่าเกณฑ์ 199 ค่า ซึ่งอยู่ในเงื่อนไข คือ ข้อมูลต้องอยู่ในช่วง 150-200 ค่า ดังนั้นจึงเลือกตัดค่าเกินเกณฑ์ที่ 80 ดังรูปภาพที่ 4-4

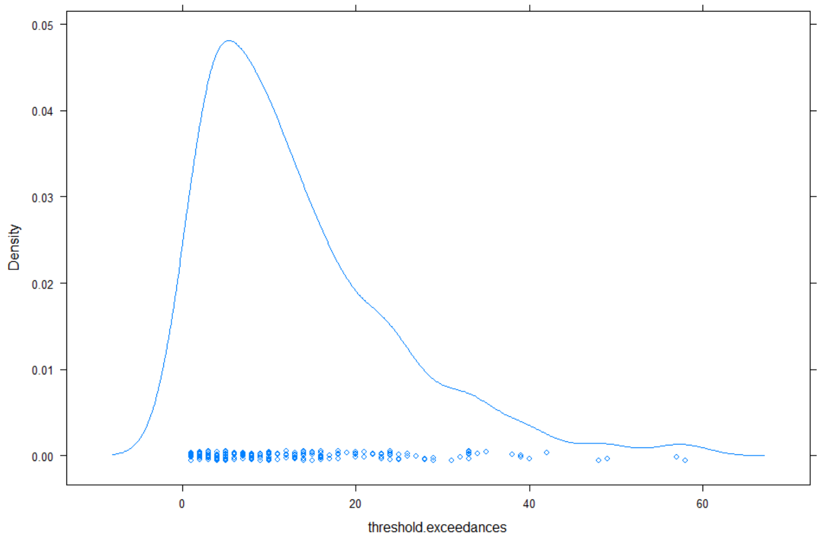


รูปภาพ 4-3 กราฟแสดงค่าความคลาดเคลื่อนเฉลี่ย (mean residual life plot : MRL) ของสถานี สถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลาง



รูปภาพ 4-4 กราฟแสดงการตัดค่าเกินเกณฑ์ที่ 80

4. นำข้อมูลที่เหลือจากการตัดค่าเกินเกณฑ์จำนวน 199 ข้อมูลมาแสดงกราฟฟังก์ชันการแจกแจงความน่าจะเป็นดังรูปภาพที่ 4-5



รูปภาพ 4-5 กราฟแสดงฟังก์ชันการแจกแจงความน่าจะเป็นของค่าเกินเกณฑ์

4.1.1.2 ประมาณค่าพารามิเตอร์ที่เหมาะสม

หาตัวแบบจำลองที่เหมาะสมสำหรับข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของเขตวังทองหลางด้วยวิธีการแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution: GPD) แสดงดังตารางที่ 4-4

ตารางที่ 4-4 การประมาณค่าพารามิเตอร์ที่เหมาะสมของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ของสถานีสถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลาง

ค่าพารามิเตอร์	AIC	BIC
$\hat{\sigma}$ =15.6139392, $\hat{\xi}$ = 0.1787775	-1416.525	-1409.938

ประมาณค่าพารามิเตอร์ มีค่าเท่ากับ 0.1787775 (0.057776) เมื่อประมาณช่วงความเชื่อมั่นจากค่าความคลาดเคลื่อนมาตรฐานจะได้ $\hat{\xi}(s.e)$

$$0.1787775 \pm 1.96(0.057776) = (-0.06554, -0.29202)$$

เมื่อสร้างแบบจำลองค่าขีดสุดซึ่งวิเคราะห์ข้อมูลด้วยวิธีค่าเกินเกณฑ์โดยประมาณค่าพารามิเตอร์ด้วยวิธี MLE พบว่า ช่วงความเชื่อมั่นอยู่ระหว่าง -0.06554 ถึง -0.29202 จะเห็นได้ว่า ช่วงความเชื่อมั่นของ $\hat{\xi}$ ครอบคลุมศูนย์ $(\hat{\xi} < 0)$ ดังนั้น ตัวแบบจำลองนี้มีการแจกแจงพาเรโต II

4.1.1.3 ตรวจสอบแบบจำลอง

การตรวจสอบความเหมาะสมการแจกแจงของตัวแบบจำลองโดยใช้สถิติทดสอบ The Komolgorov-simrnov test (KS-Test) ของสถานีสถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลาง ว่ามีการแจกแจงพาเรโต II หรือไม่

สมมติฐานทดสอบ

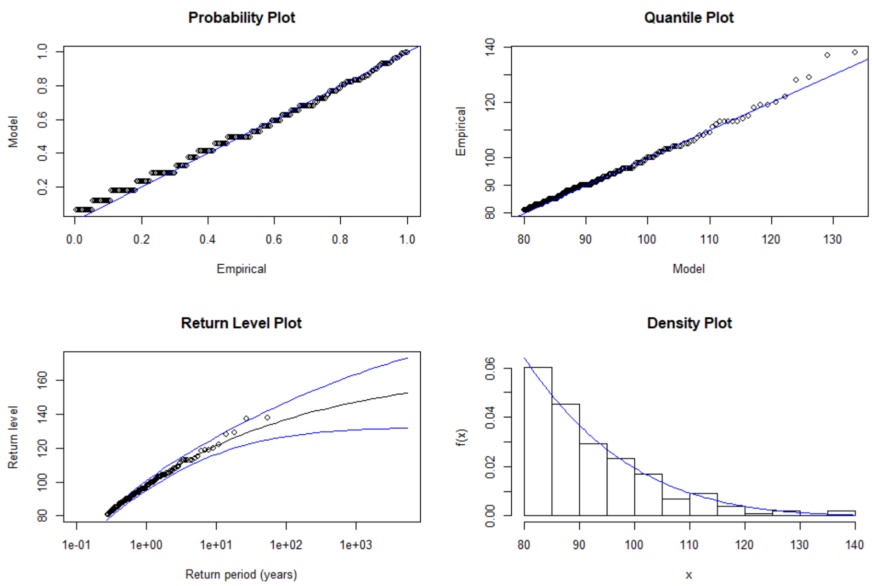
H_0 : มีการแจกแจงพาเรโต II

H_1 : ไม่มีการแจกแจงพาเรโต II

ระดับนัยสำคัญที่ 0.05

เนื่องจาก ค่า P-value เท่ากับ 0.2527 มากกว่า ระนัยสำคัญ 0.05 จึงทำการยอมรับสมมติฐานหลัก ดังนั้น สถานีสถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลาง มีการแจกแจงพาเรโต II เป็นการแจกแจงที่เหมาะสมตัวแบบจำลอง

นอกจากนี้ยังสามารถใช้กราฟความน่าจะเป็น (Probability Plot) และกราฟควอนไทล์ (Quantile Plot) เพื่อตรวจสอบความเหมาะสมของตัวแบบจำลองดังที่ได้แสดงในภาพที่ 4-6



รูปภาพ 4-6 กราฟแสดงฟังก์ชันการแจกแจงพาเรโตวางนัยทั่วไป (GPD) ของสถานีสถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลาง

4.1.1.4 การประมาณระดับการเกิดซ้ำ

จากตัวแบบจำลองที่ได้สามารถประมาณระดับการเกิดซ้ำของปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของเขตวังทองหลางในอีก 2, 5, 10, 50 และ 100 ปี ข้างหน้า (หน่วย ไมโครกรัม/ลูกบาศก์เมตร: วัน) ได้ดังตารางที่ 4-5

ตารางที่ 4-5 แสดงค่าประมาณระดับการเกิดซ้ำ ณ ช่วงเวลาที่เกิดซ้ำต่าง ๆ

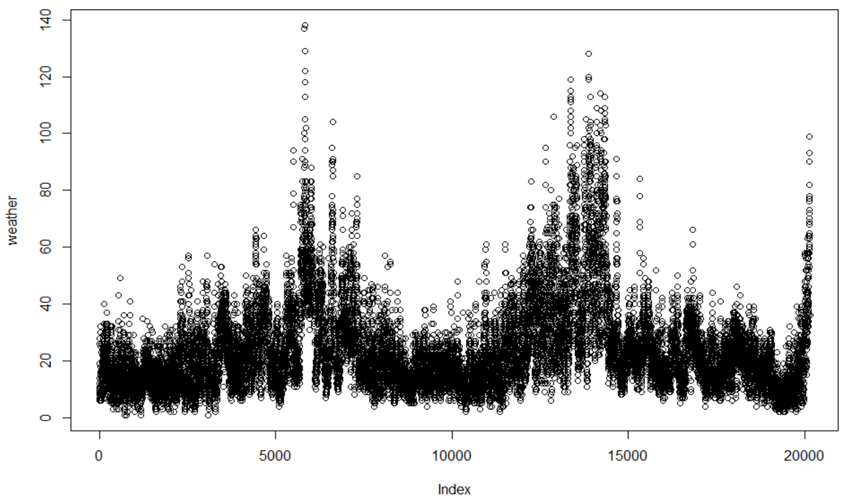
2-year level	5-year level	10-year level	50-year level	100-year level
105.9907	115.2559	121.3220	132.8173	136.8352

จากตารางที่ 4-5 พบว่า เมื่อรอบปีการเกิดซ้ำเพิ่มขึ้นจะทำให้ค่าระดับการเกิดซ้ำเพิ่มขึ้นตามไปด้วย โดยตัวแบบจำลองนี้สามารถใช้คาดการณ์ระดับการเกิดซ้ำของปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของสถานีสถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลาง ในรอบปีที่ต้องการทราบได้ ทำให้สามารถนำไปปรับใช้วางแผนควบคุมปริมาณความเข้มข้นของ $PM_{2.5}$ เพื่อคุณภาพชีวิตที่ดีของประชาชน

4.1.2 สถานีการเคหะดินแดง เขตดินแดง

4.1.2.1 การเตรียมข้อมูล

1. นำข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของสถานีการเคหะดินแดง เขตดินแดง นำมาแสดงผลดังรูปภาพที่ 4-7



รูปภาพ 4-7 กราฟแสดงการกระจายของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของสถานีการเคหะดินแดง เขตดินแดง

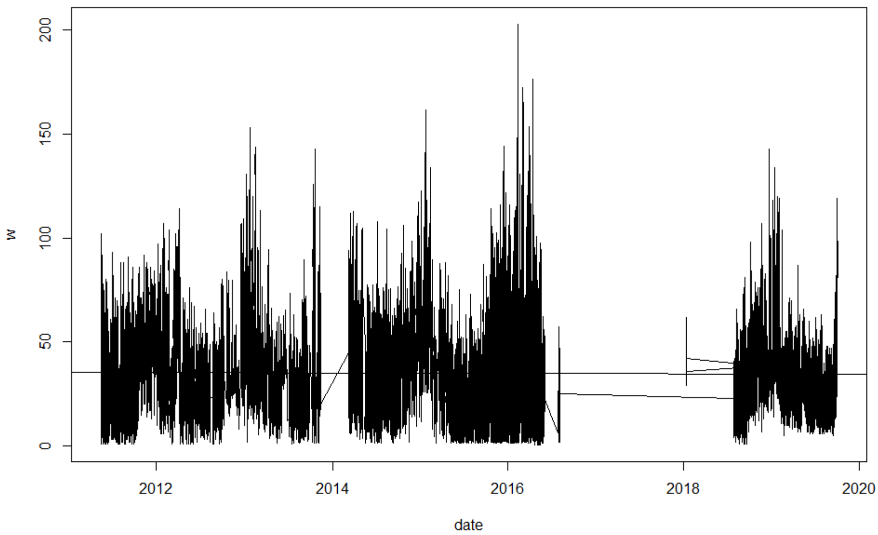
2. การทดสอบความคงที่ (Stationary) และแนวโน้ม (Trend) ของข้อมูล

ตารางที่ 4-6 การทดสอบความคงที่และแนวโน้มของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ของสถานีการเคหะดินแดง เขตดินแดง

การทดสอบ	ค่าสถิติ	P-value
Augmented Dickey-Fuller (ADF) test	-14.514	0.01
Mann-Kendall trend test	-3.6425	0.00027

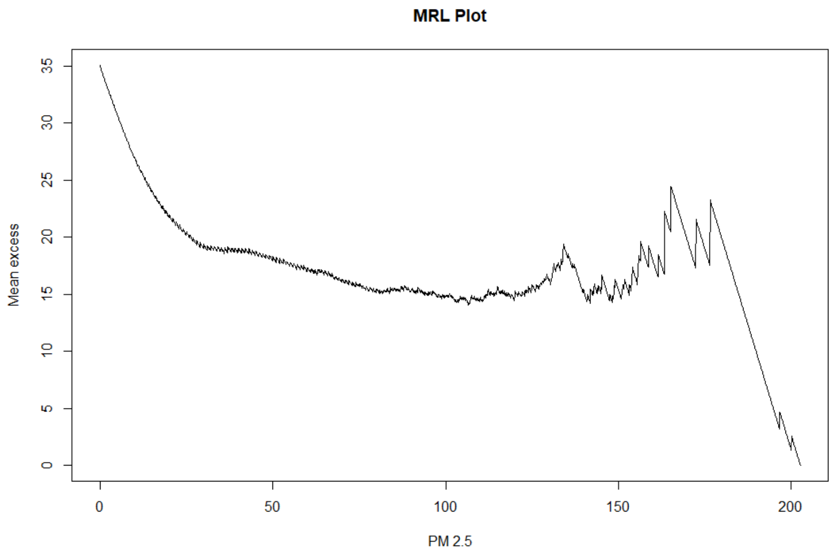
*(*ระดับนัยสำคัญ 0.05)*

จากตารางที่ 4-6 แสดงการทดสอบ จากวิธีการทดสอบของ Augmented Dickey-Fuller (ADF) test ที่ระดับนัยสำคัญ 0.05 เมื่อพิจารณาค่า P-value เท่ากับ 0.01 พบว่าข้อมูลชุดนี้อยู่ภายใต้กระบวนการคงที่ และพิจารณาจากวิธีทดสอบของ Mann-Kendall trend (MK) test ที่ระดับนัยสำคัญ 0.05 เมื่อพิจารณาค่า P-value เท่ากับ 0.00027 พบว่าข้อมูลชุดนี้ไม่มีแนวโน้มของข้อมูล แสดงดังรูปภาพที่ 4-8 ดังนี้

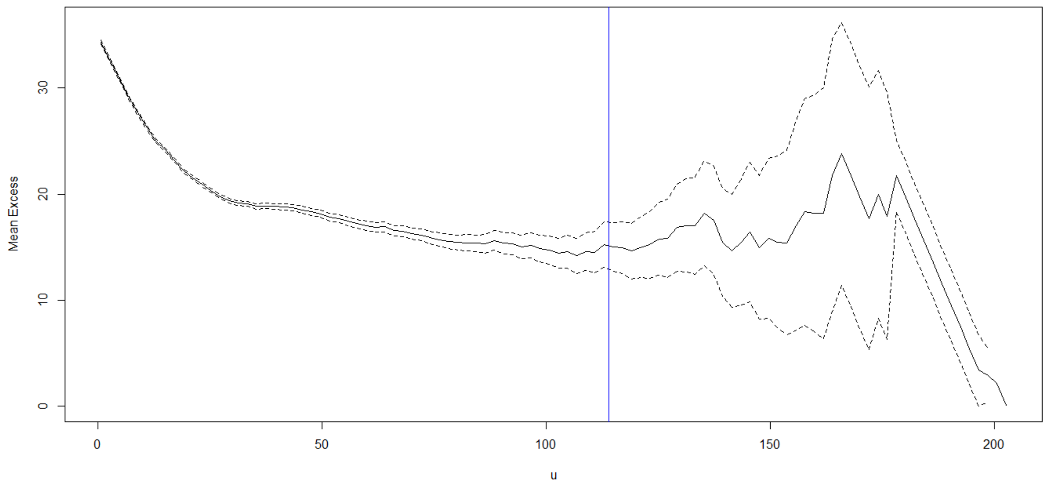


รูปภาพ 4-8 กราฟแสดง Time series ของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของสถานีการเคหะดินแดง เขตดินแดง

3. ทำการเลือกค่าเกินเกณฑ์ (μ_0) จากกราฟค่าความคลาดเคลื่อนเฉลี่ย (mean residual life plot: MRL) ดังรูปภาพที่ 4-9 โดยกำหนดค่าเกณฑ์ที่เปอร์เซ็นไทล์ 99.6% มีค่าเท่ากับ 114 เนื่องจากมีจำนวนข้อมูลที่ค่าเกินค่าเกณฑ์ 189 ค่า ซึ่งอยู่ในเงื่อนไข คือ ข้อมูลต้องอยู่ในช่วง 150-200 ค่า ดังนั้นจึงเลือกตัดค่าเกินเกณฑ์ที่ 114 ดังรูปภาพที่ 4-10

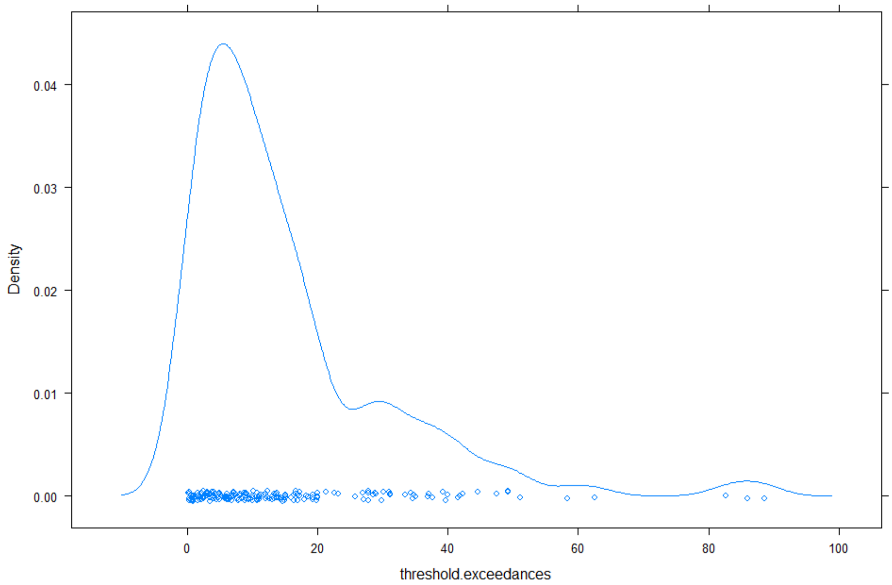


รูปภาพ 4-9 กราฟแสดงค่าความคลาดเคลื่อนเฉลี่ย (mean residual life plot : MRL) ของสถานีการเคหะดินแดง เขตดินแดง



รูปภาพ 4-10 กราฟแสดงการตัดค่าเกินเกณฑ์ที่ 114

4. นำข้อมูลที่เหลือจากการตัดค่าเกินเกณฑ์จำนวน 189 ข้อมูลมาแสดงกราฟฟังก์ชันการแจกแจงความน่าจะเป็นดังรูปภาพที่ 4-11



รูปภาพ 4-11 กราฟแสดงฟังก์ชันการแจกแจงความน่าจะเป็นของค่าเกินเกณฑ์

4.1.2.2 ประมาณค่าพารามิเตอร์ที่เหมาะสม

หาตัวแบบจำลองที่เหมาะสมสำหรับข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของเขตดินแดงด้วยวิธีการแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution : GPD) แสดงดังตารางที่ 4-7

ตารางที่ 4-7 การประมาณค่าพารามิเตอร์ที่เหมาะสมของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ของสถานีการเคหะดินแดง เขตดินแดง

ค่าพารามิเตอร์	AIC	BIC
$\hat{\sigma}$ =14.63642754, $\hat{\xi}$ =0.03388451	-1401.122	-1394.638

ประมาณค่าพารามิเตอร์ มีค่าเท่ากับ -0.03388451(0.07646826) เมื่อประมาณช่วงความเชื่อมั่นจากค่าความคลาดเคลื่อนมาตรฐานจะได้ $\hat{\xi}$(s.e)

$$-0.03388451 \pm 1.96(0.07646826) = (-0.18376, 0.11599)$$

เมื่อสร้างแบบจำลองค่าขีดสุดซึ่งวิเคราะห์ข้อมูลด้วยวิธีค่าเกินเกณฑ์โดยประมาณค่าพารามิเตอร์ด้วยวิธี MLE พบว่า ช่วงความเชื่อมั่นอยู่ระหว่าง -0.18376 ถึง 0.11599 จะเห็นได้ว่าช่วงความเชื่อมั่นของ ξ ครอบคลุมศูนย์ $(\hat{\xi} \to 0)$ ดังนั้น ตัวแบบจำลองนี้มีการแจกแจงเลขชี้กำลัง

4.1.2.3 ตรวจสอบแบบจำลอง

การตรวจสอบความเหมาะสมการแจกแจงของตัวแบบจำลองโดยใช้สถิติทดสอบ The Komolgorov-simrnov test (KS-Test) ของสถานีการเคหะดินแดง เขตดินแดง มีการแจกแจงเลขชี้กำลัง หรือไม่

สมมติฐานทดสอบ

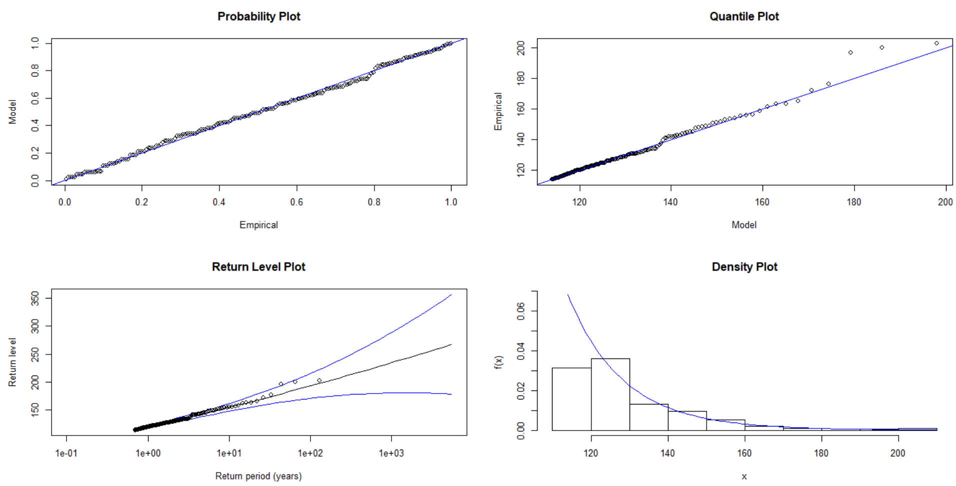
H_0 : มีการแจกแจงเลขชี้กำลัง

H_1 : ไม่มีการแจกแจงเลขชี้กำลัง

ระดับนัยสำคัญที่ 0.05

เนื่องจาก ค่า P-value เท่ากับ 0.832 มากกว่า ระดับนัยสำคัญที่ 0.05 จึงทำการยอมรับสมมติฐานหลัก ดังนั้น สถานีการเคหะดินแดง เขตดินแดง มีการแจกแจงเลขชี้กำลังเป็นการแจกแจงที่เหมาะสมของตัวแบบจำลอง

นอกจากนี้ยังสามารถใช้กราฟความน่าจะเป็น (Probability Plot) และกราฟควอนไทล์ (Quantile Plot) เพื่อตรวจสอบความเหมาะสมของตัวแบบจำลองดังที่ได้แสดงในรูปภาพที่ 4-12



รูปภาพ 4-12 กราฟแสดงฟังก์ชันการแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution: GPD) ของสถานีการเคหะดินแดง เขตดินแดง

4.1.2.4 การประมาณระดับการเกิดซ้ำ

จากตัวแบบจำลองที่ได้ สามารถประมาณระดับการเกิดซ้ำของปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของเขตดินแดงในอีก 2, 5, 10, 50 และ 100 ปี ข้างหน้า (หน่วย ไมโครกรัม/ลูกบาศก์เมตร: วัน) ได้ดังตารางที่ 4-8

ตารางที่ 4-8 แสดงค่าประมาณระดับการเกิดซ้ำ ณ ช่วงเวลาที่เกิดซ้ำต่าง ๆ

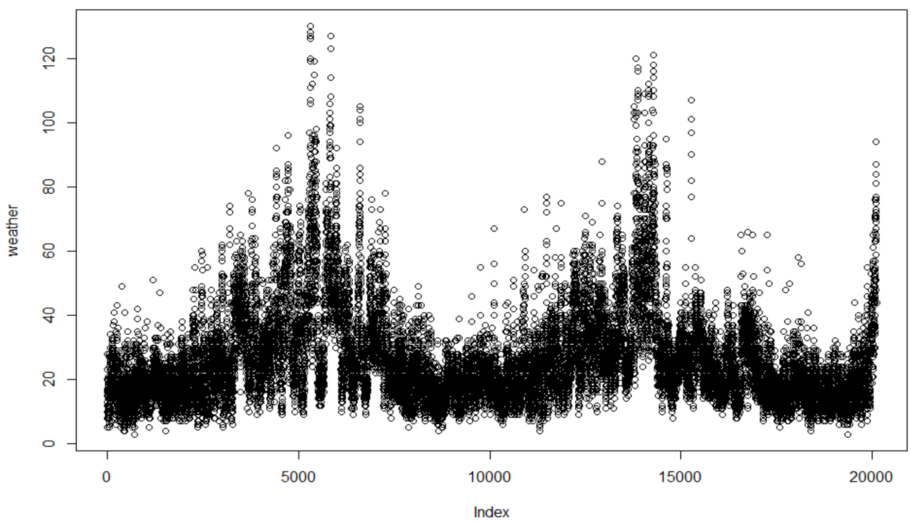
2-year level	5-year level	10-year level	50-year level	100-year level
129.7639	143.8803	154.8572	181.3702	193.2460

จากตารางที่ 4-8 พบว่า เมื่อรอบปีการเกิดซ้ำเพิ่มขึ้นจะทำให้ค่าระดับการเกิดซ้ำเพิ่มขึ้นตามไปด้วย โดยตัวแบบจำลองนี้สามารถใช้คาดการณ์ระดับการเกิดซ้ำของปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของสถานีการเคหะดินแดง เขตดินแดง ในรอบปีที่ต้องการทราบได้ ทำให้สามารถนำไปปรับใช้วางแผนควบคุมปริมาณความเข้มข้นของ $PM_{2.5}$ เพื่อคุณภาพชีวิตที่ดีของประชาชน

4.1.3 สถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน

4.1.3.1 การเตรียมข้อมูล

1. นำข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวันนำมาแสดงผลดังภาพที่ 4-13



รูปภาพ 4-13 กราฟแสดงการกระจายของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน

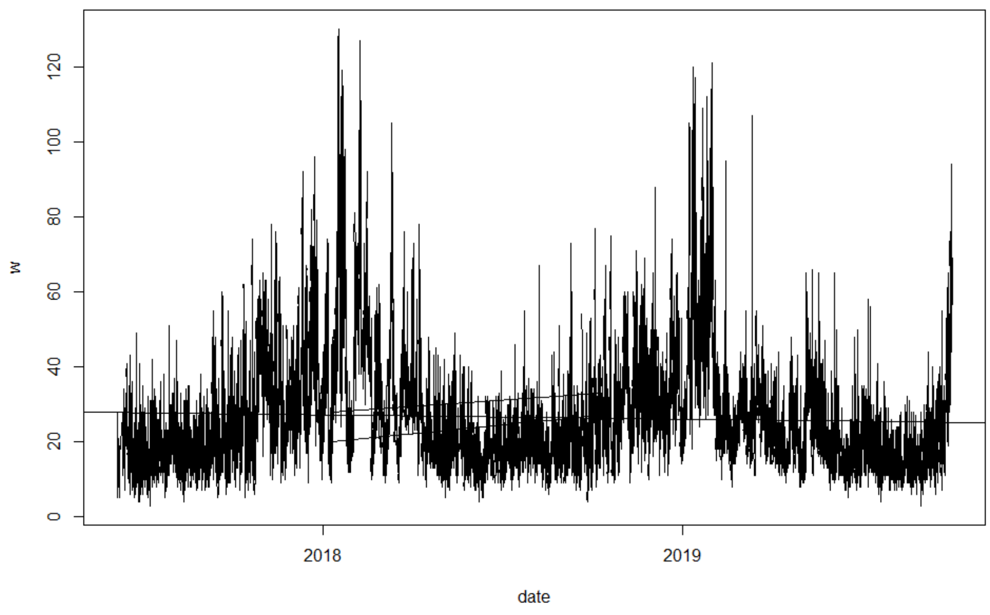
2. การทดสอบความคงที่ (Stationary) และแนวโน้ม (Trend) ของข้อมูล

ตารางที่ 4-9 การทดสอบความคงที่และแนวโน้มของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ของสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน

การทดสอบ	ค่าสถิติ	P-value
Augmented Dickey-Fuller (ADF) test	-9.7604	0.01
Mann-Kendall trend test	-6.4776	9.322e-11

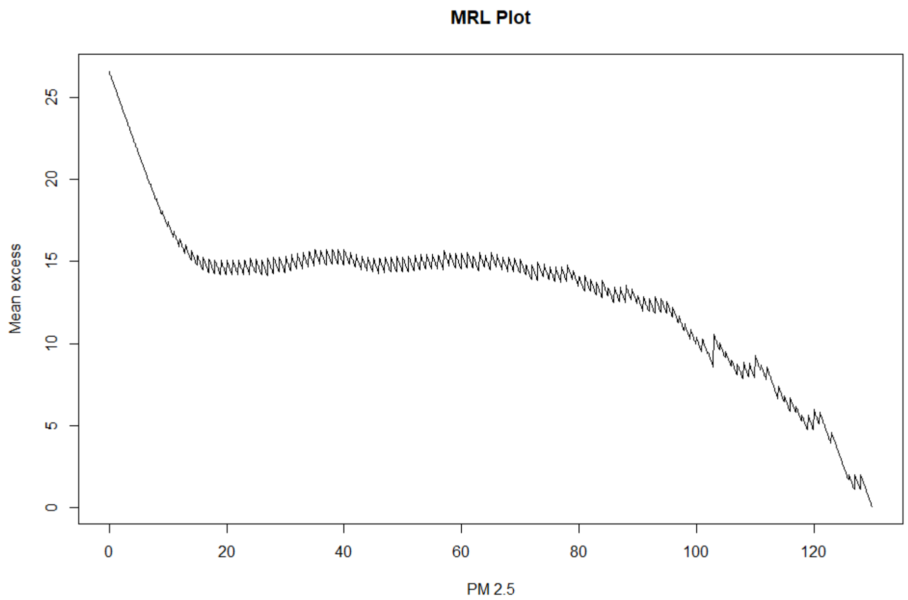
*(*ระดับนัยสำคัญ 0.05)*

จากตารางที่ 4-9 แสดงการทดสอบ จากวิธีการทดสอบของ Augmented Dickey-Fuller (ADF) test ที่ระดับนัยสำคัญ 0.05 เมื่อพิจารณาค่า P-value เท่ากับ 0.01 พบว่าข้อมูลชุดนี้อยู่ภายใต้กระบวนการคงที่ และพิจารณาจากวิธีทดสอบของ Mann-Kendall trend (MK) test ที่ระดับนัยสำคัญ 0.05 เมื่อพิจารณาค่า P-value เท่ากับ 9.322e-11 พบว่าข้อมูลชุดนี้ไม่มีแนวโน้มของข้อมูลดังรูปภาพที่ 4-14 ดังนี้

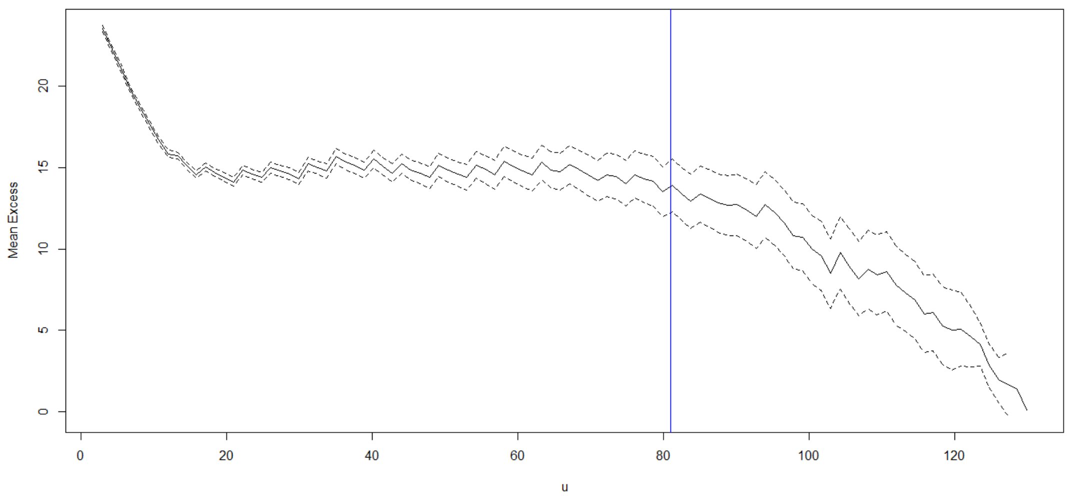


รูปภาพ 4-14 กราฟแสดง Time series ของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน

3. ทำการเลือกค่าเกินเกณฑ์ (μ_0)จากกราฟค่าความคลาดเคลื่อนเฉลี่ย (mean residual life plot : MRL) ดังรูปภาพที่ 4-15 โดยกำหนดค่าเกณฑ์ที่เปอร์เซ็นไทล์ 99% มีค่าเท่ากับ 81 เนื่องจากมีจำนวนข้อมูลที่ค่าเกินค่าเกณฑ์ 196 ค่า ซึ่งอยู่ในเงื่อนไข คือ ข้อมูลต้องอยู่ในช่วง 150-200 ค่า ดังนั้นจึงเลือกตัดค่าเกินเกณฑ์ที่ 81 ดังรูปภาพที่ 4-16

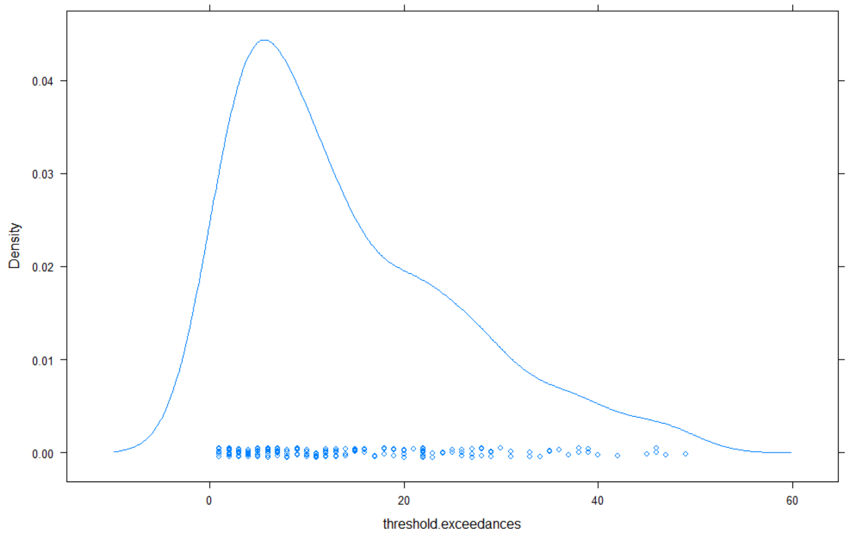


รูปภาพ 4-15 กราฟแสดงค่าความคลาดเคลื่อนเฉลี่ย (mean residual life plot : MRL) ของสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน



รูปภาพ 4-16 กราฟแสดงการตัดค่าเกินเกณฑ์ที่ 81

4. นำข้อมูลที่เหลือจากการตัดค่าเกินเกณฑ์จำนวน 196 ข้อมูลมาแสดงกราฟฟังก์ชันการแจกแจงความน่าจะเป็นดังรูปภาพที่ 4-17



รูปภาพ 4-17 กราฟแสดงฟังก์ชันการแจกแจงความน่าจะเป็นของค่าเกินเกณฑ์

4.1.3.2 ประมาณค่าพารามิเตอร์ที่เหมาะสม

หาตัวแบบจำลองที่เหมาะสมสำหรับข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของเขตปทุมวันด้วยวิธีการแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution : GPD) แสดงดังตารางที่ 4-10

ตารางที่ 4-10 การประมาณค่าพารามิเตอร์ที่เหมาะสมของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ของสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน

ค่าพารามิเตอร์	AIC	BIC
$\hat{\sigma}$ =18.4287776, $\hat{\xi}$ =-0.2983433	-1413.355	-1406.798

ประมาณค่าพารามิเตอร์ มีค่าเท่ากับ -0.2983433(0.06889261) เมื่อประมาณช่วงความเชื่อมั่นจากค่าความคลาดเคลื่อนมาตรฐานจะได้ $\hat{\xi}(s.e)$

-0.2983433 ± 1.96(0.06889261) = (-0.43337, -0.16331)

เมื่อสร้างแบบจำลองค่าขีดสุดซึ่งวิเคราะห์ข้อมูลด้วยวิธีค่าเกินเกณฑ์โดยประมาณค่าพารามิเตอร์ด้วยวิธี MLE พบว่า ช่วงความเชื่อมั่นอยู่ระหว่าง -0.43337 ถึง -0.16331 จะเห็นได้ว่าช่วงความเชื่อมั่นของ $\hat{\xi}$ ครอบคลุมศูนย์ ($\hat{\xi} < 0$) ดังนั้น ตัวแบบจำลองนี้มีการแจกแจงพาเรโต II

4.1.3.2 ตรวจสอบแบบจำลอง

การตรวจสอบความเหมาะสมการแจกแจงของตัวแบบจำลองโดยใช้สถิติทดสอบ The Komolgorov-simrnov Test (KS-Test) ของสถานีการเคหะดินแดง เขตดินแดงมีการแจกแจงพาเรโต II หรือไม่

สมมติฐานทดสอบ

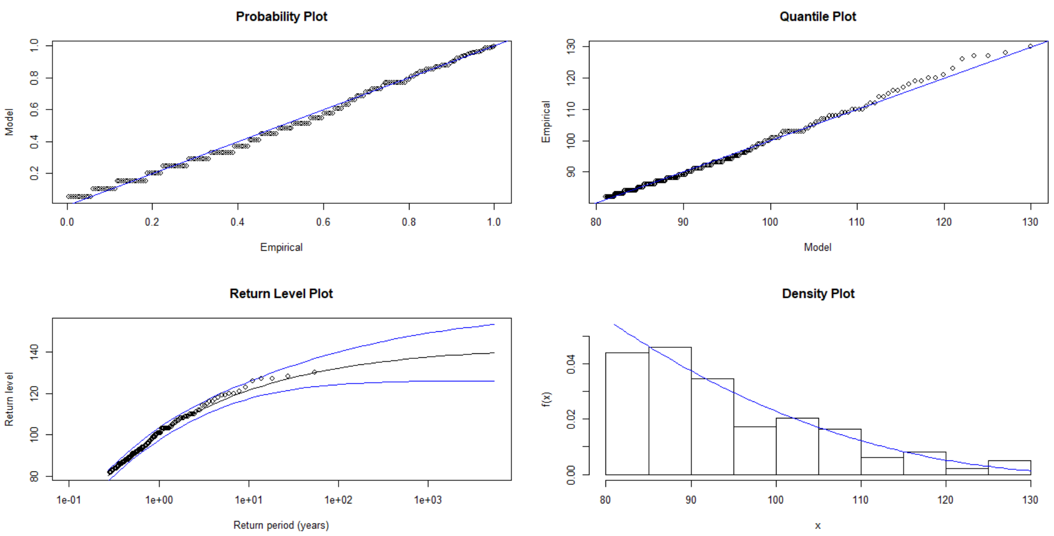
H_0 : มีการแจกแจงพาเรโต II

H_1 : ไม่มีการแจกแจงพาเรโต II

ระดับนัยสำคัญที่ 0.05

เนื่องจาก ค่า P-value เท่ากับ 0.571 มากกว่า ระดับนัยสำคัญที่ 0.05 จึงทำการยอมรับสมมติฐานหลัก ดังนั้น สถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน มีการแจกแจงพาเรโต II เป็นการแจกแจงที่เหมาะสมของตัวแบบจำลอง

นอกจากนี้ยังสามารถใช้กราฟความน่าจะเป็น (Probability Plot) และกราฟควอนไทล์ (Quantile Plot) เพื่อตรวจสอบความเหมาะสมของตัวแบบจำลองดังที่ได้แสดงในภาพที่ 4-18



รูปภาพ 4-18 กราฟแสดงฟังก์ชันการแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution: GPD) ของสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน

4.1.3.3 การประมาณระดับการเกิดซ้ำ

จากตัวแบบจำลองที่ได้ สามารถประมาณระดับการเกิดซ้ำของปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของเขตวังทองหลางในอีก 2, 5, 10, 50 และ 100 ปี ข้างหน้า (หน่วย ไมโครกรัม/ลูกบาศก์เมตร : วัน) ได้ดังตารางที่ 4-5

ตารางที่ 4-11 แสดงค่าประมาณระดับการเกิดซ้ำ ณ ช่วงเวลาที่เกิดซ้ำต่าง ๆ

2-year level	5-year level	10-year level	50-year level	100-year level
108.3851	116.6103	121.4974	129.6083	132.0663

จากตารางที่ 4-11 พบว่า เมื่อรอบปีการเกิดซ้ำเพิ่มขึ้นจะทำให้ค่าระดับการเกิดซ้ำเพิ่มขึ้นตามไปด้วย โดยตัวแบบจำลองนี้สามารถใช้คาดการณ์ระดับการเกิดซ้ำของปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงของสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน ในรอบปีที่ต้องการทราบได้ ทำให้สามารถนำไปปรับใช้วางแผนควบคุมปริมาณความเข้มข้นของ $PM_{2.5}$ เพื่อคุณภาพชีวิตที่ดีของประชาชน

4.2 ตัวแบบจำลองการแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) จากปริมาณความเข้มข้นของ PM2.5 รายสัปดาห์ของ 3 เขตธุรกิจในกรุงเทพมหานคร

จากผลการวิเคราะห์เพื่อหาแบบจำลองที่เหมาะสมของปริมาณความเข้มข้นของ $PM_{2.5}$ ใน 3 เขตของกรุงเทพมหานคร โดยมีข้อมูลพื้นฐานรายสัปดาห์ผลสรุปดังตารางที่ 4-12

ตารางที่ 4-12 ข้อมูลรายสัปดาห์พื้นฐานของปริมาณความเข้มข้นของ $PM_{2.5}$ ใน 3 เขตธุรกิจในกรุงเทพมหานคร

สถานี	จำนวนข้อมูล	ค่าสูงสุด	ค่าเฉลี่ย	S. D	ค่าต่ำสุด
สถานีสถานีตำรวจนครบาลโชคชัย เขตวังทองหลาง	122	138	53.64	29.07	19
สถานีการเคหะดินแดง เขตดินแดง	312	202.64	79.03	26.87	37
สถานีโรงพยาบาลจุฬาลงกรณ์เขตปทุมวัน	122	130	57.79	23.62	28

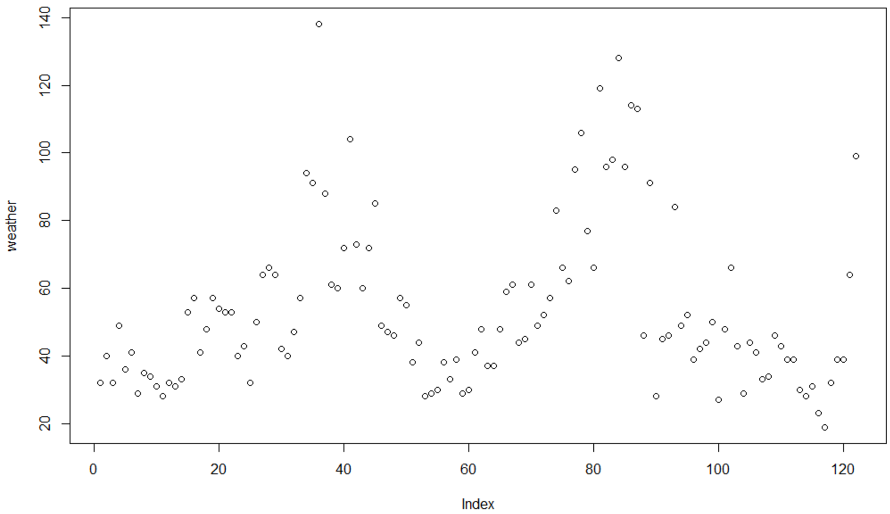
(ไมโครกรัม/ลูกบาศก์เมตร (μg/m3))

4.1.1 สถานีสถานีตำรวจนครบาลโชคชัย เขตวังทองหลาง

4.1.1.1 การเตรียมข้อมูล

1. นำข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของสถานีสถานีตำรวจนครบาลโชคชัย เขตวังทองหลาง แสดงดังรูปภาพที่ 4-19

]



รูปภาพ 4-19 กราฟแสดงการกระจายของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน

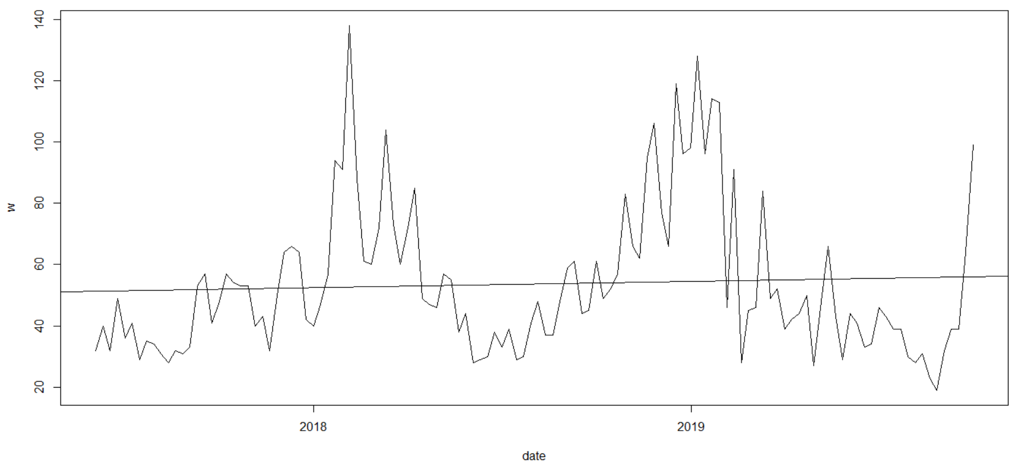
2. การทดสอบความคงที่ (Stationary) และแนวโน้ม (Trend) ของข้อมูล

ตารางที่ 4-13 การทดสอบความคงที่และแนวโน้มของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของสถานีสถานีตำรวจนครบาลโชคชัย เขตวังทองหลาง

การทดสอบ	ค่าสถิติ	P-value
Augmented Dickey-Fuller (ADF) test	-2.538	0.353
Mann-Kendall trend test	0.21474	0.83

*(*ระดับนัยสำคัญ 0.05)*

จากตารางที่ 4-13 แสดงการทดสอบ จากวิธีการทดสอบของ Augmented Dickey-Fuller (ADF) test ที่ระดับนัยสำคัญ 0.05 เมื่อพิจารณาค่า P-value เท่ากับ 0.353 พบว่า ข้อมูลชุดนี้ไม่อยู่ภายใต้กระบวนการคงที่และพิจารณาจากวิธีทดสอบของ Mann-Kendall trend (MK) test ที่ระดับนัยสำคัญ 0.05 เมื่อพิจารณาค่า P-value เท่ากับ 0.83 พบว่าข้อมูลชุดนี้มีแนวโน้มของข้อมูลแสดงดังภาพที่ 4-20 ดังนี้



รูปภาพ 4-20 กราฟแสดง Time series ของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของสถานีสถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลาง

4.1.1.2 ประมาณค่าพารามิเตอร์ที่เหมาะสม

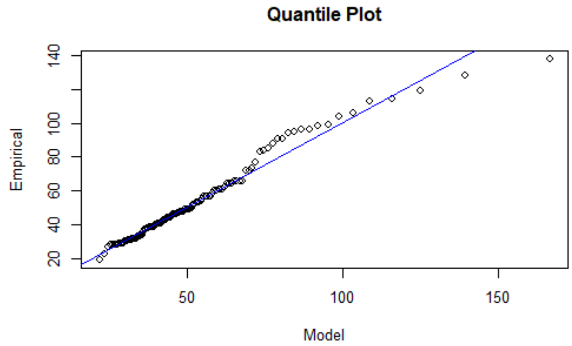
การหาตัวแบบจำลองที่เหมาะสมที่สุดสำหรับข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ ทำได้โดยการหาค่าประมาณพารามิเตอร์ ช่วงความเชื่อมั่น และการแจกแจงของตัวแบบจำลอง จากวิธีการแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) แล้วคัดเลือกตัวแบบที่เหม

าะสมที่สุดโดยใช้เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike’s information criterion: AIC) และเกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criteria: BIC) จากแบบจำลอง 4 รูปแบบแสดงดังตารางที่ 4-14

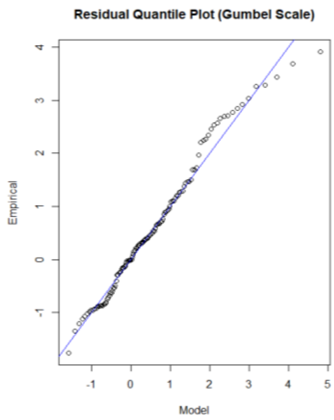
ตารางที่ 4-14 การประมาณค่าพารามิเตอร์ที่เหมาะสมของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ของสถานีสถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลาง

รูปแบบ	ค่าพารามิเตอร์	AIC	BIC
1	μ =41.444, $\hat{\sigma}$ =14.662, $\hat{\xi}$ =0.220	-1065.656	-1057.244
2	$\mu(t)$ =43.256-0.034t, $\hat{\sigma}$ =14.300, $\hat{\xi}$ =0.255	-1062.085	-1050.869
3	$\mu(t)$ =43.230-0.001t-0.0004t^2, $\hat{\sigma}$ =14.009, $\hat{\xi}$ =0.275	-1057.524	-1043.504
4	$\mu(t)$ = 38.540+0.048t, $\hat{\sigma}(t) = \exp$ (2.241+ 0.0063t), $\hat{\xi}$ =0.278	-1055.883	-1041.863
5	-	-	-

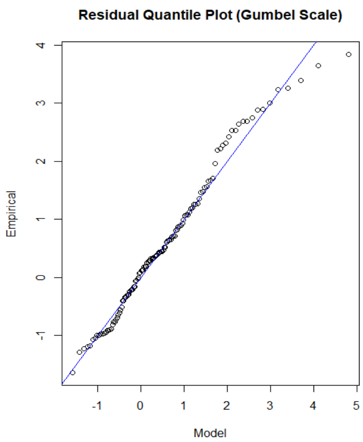
จากภาพที่ 4-21 ถ้าตัวแบบจำลองมีการแจกแจงที่เหมาะสม Q-Q plots จะอยู่ใกล้เส้นทแยงมุมซึ่งกราฟจะเป็นเส้นตรง พบว่า ตัวแบบจำลองที่ 1-4 มีการแจกแจงที่เหมาะสม เนื่องจาก Q-Q plots เกาะกลุ่มอยู่ใกล้กับเส้นทแยงมุมและกราฟเป็นเส้นตรง ส่วนตัวแบบจำลองที่ 5 มีการแจกแจงที่ไม่เหมาะสม เนื่องจาก Q-Q plots ไม่อยู่ใกล้กับเส้นทแยงมุมและกราฟไม่เป็นเส้นตรง ดังนั้นตัวแบบจำลองที่จะนำไปพิจารณาหาการแจกแจงที่เหมาะสมที่สุด คือ ตัวแบบจำลองที่ 1-4 และจากตารางที่ 4-14 เมื่อพิจารณาค่า AIC และ BIC พบว่า ตัวแบบที่ 1 เป็นตัวแบบจำลองที่ให้ค่า AIC และ BIC น้อยที่สุด นั้นคือ เป็นตัวแบบที่เหมาะสมที่สุดสำหรับข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ของสถานีสถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลาง ดังนั้น สามารถเขียนอธิบายได้ว่า $x \sim GEV(\mu,\sigma,\xi)$ โดยที่พารามิเตอร์ μ =41.444, $\hat{\sigma}$ =14.662 และ $\hat{\xi}$ =0.220 มีค่าคงที่



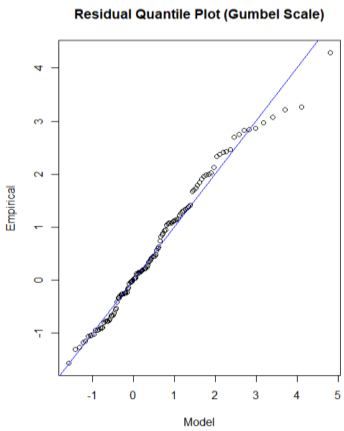
ตัวแบบที่ 1



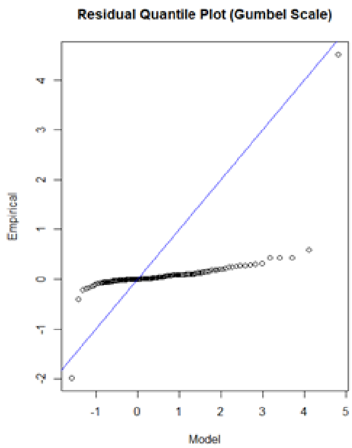
ตัวแบบที่ 2



ตัวแบบที่ 3



ตัวแบบที่ 4



ตัวแบบที่ 5

รูปภาพ 4-21 กราฟ Q-Q Plot ของตัวแบบ

ประมาณค่าพารามิเตอร์ มีค่าเท่ากับ 0.220(0.0788) เมื่อประมาณช่วงความเชื่อมั่นจากค่าความคลาดเคลื่อนมาตรฐานจะได้ $\hat{\xi}(\text{s.e})$

$$0.220 \pm 1.96(0.0788) = (0.0656, 0.3745)$$

เมื่อสร้างแบบจำลองค่าขีดสุดซึ่งวิเคราะห์ข้อมูลด้วยวิธีค่าเกินเกณฑ์โดยประมาณค่าพารามิเตอร์ด้วยวิธี MLE พบว่า ช่วงความเชื่อมั่นอยู่ระหว่าง 0.0656 ถึง 0.3745 จะเห็นได้ว่าช่วงความเชื่อมั่นของ $\hat{\xi}$ มากกว่าศูนย์ $(\hat{\xi} > 0)$ ดังนั้น ตัวแบบจำลองนี้มีการแจกแจงฟรีเซท

4.1.1.3 ตรวจสอบแบบจำลอง

การตรวจสอบความเหมาะสมการแจกแจงของตัวแบบจำลองโดยใช้สถิติทดสอบ The Komolgorov-simrnov test (KS-Test) ของสถานีสถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลางมีการแจกแจงฟรีเซท หรือไม่

สมมติฐานทดสอบ

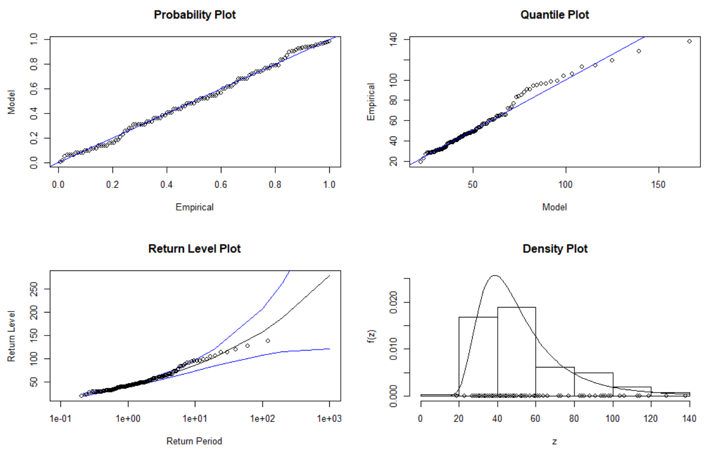
H_0 : มีการแจกแจงฟรีเซท

H_1 : ไม่มีการแจกแจงฟรีเซท

ระดับนัยสำคัญที่ 0.05

เนื่องจาก ค่า P-value เท่ากับ 0.9071 มากกว่า ระนัยสำคัญ 0.05 จึงทำการยอมรับสมมติฐานหลัก ดังนั้น สถานีสถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลางมีการแจกแจงฟรีเซทเป็นการแจกแจงที่เหมาะสมของตัวแบบจำลอง

นอกจากนี้ยังสามารถใช้กราฟความน่าจะเป็น (Probability Plot) และกราฟควอนไทล์ (Quantile Plot) เพื่อตรวจสอบความเหมาะสมของตัวแบบจำลองดังที่ได้แสดงในรูปภาพที่ 4-21



รูปภาพ 4-22 กราฟแสดงฟังก์ชันการแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) ของสถานีสถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลาง

4.1.1.4 การประมาณระดับการเกิดซ้ำ

จากตัวแบบจำลองที่ได้ สามารถประมาณระดับการเกิดซ้ำของปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของเขตวังทองหลางในอีก 2, 5, 10, 50 และ 100 ปี ข้างหน้า (หน่วยไมโครกรัม/ลูกบาศก์เมตร: วัน) ได้ดังตารางที่ 4-5

ตารางที่ 4-15 แสดงค่าประมาณระดับการเกิดซ้ำ ณ ช่วงเวลาที่เกิดซ้ำต่างๆ

Station	Model	2-year level	5-year level	10-year level	50-year level	100-year level
Chokchai	1	105.9907	115.2559	121.3220	132.8173	136.8352

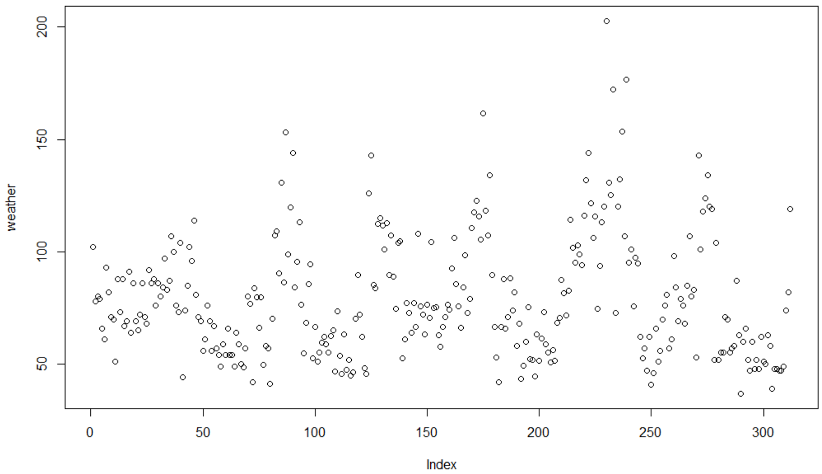
(ไมโครกรัม/ลูกบาศก์เมตร (μg/m3))

จากตารางที่ 4-15 พบว่า เมื่อรอบปีการเกิดซ้ำเพิ่มขึ้นจะทำให้ค่าระดับการเกิดซ้ำเพิ่มขึ้นตามไปด้วย โดยตัวแบบจำลองนี้สามารถใช้คาดการณ์ระดับการเกิดซ้ำของปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของสถานีสถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลาง ในรอบปีที่ต้องการทราบได้ ทำให้สามารถนำไปปรับใช้วางแผนควบคุมปริมาณความเข้มข้นของ $PM_{2.5}$ เพื่อคุณภาพชีวิตที่ดีของประชาชน

4.2.2 สถานีการเคหะดินแดง เขตดินแดง

4.2.2.1 การเตรียมข้อมูล

1. นำข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของสถานีการเคหะดินแดง เขตดินแดง แสดงผลดังรูปภาพที่ 4-23



รูปภาพ 4-23 กราฟแสดงการกระจายของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของสถานีการเคหะดินแดง เขตดินแดง

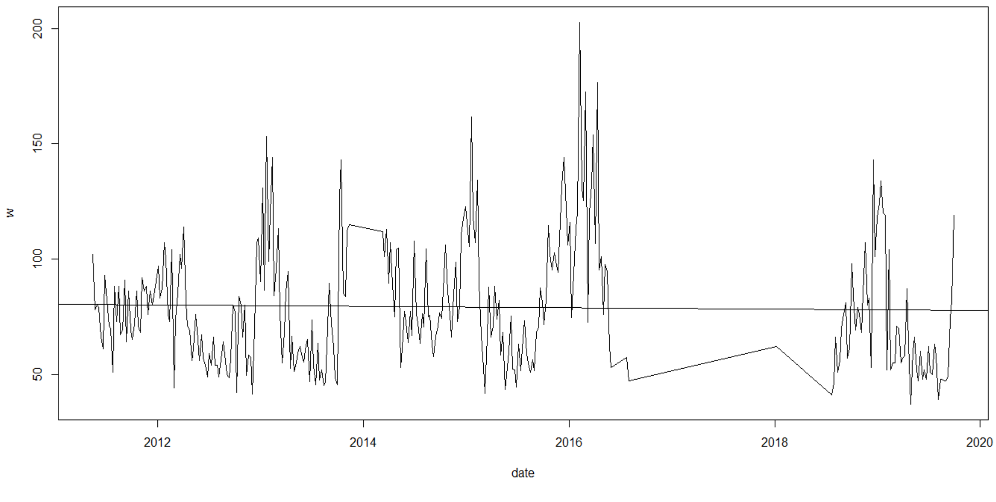
2. การทดสอบความคงที่ (Stationary) และแนวโน้ม (Trend) ของข้อมูล

ตารางที่ 4-16 การทดสอบความคงที่และแนวโน้มของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของสถานีการเคหะดินแดง เขตดินแดง

การทดสอบ	ค่าสถิติ	P-value
Augmented Dickey-Fuller (ADF) test	-4.0721	0.01
Mann-Kendall trend test	-0.8206	0.4119

*(*ระดับนัยสำคัญ 0.05)*

จากตารางที่ 4-16 แสดงการทดสอบ จากวิธีการทดสอบของ Augmented Dickey-Fuller (ADF) Test ที่ระดับนัยสำคัญ 0.05 เมื่อพิจารณาค่า P-value เท่ากับ 0.01 พบว่าข้อมูลชุดนี้อยู่ภายใต้กระบวนการคงที่ และพิจารณาจากวิธีทดสอบของ Mann-Kendall trend test ที่ระดับนัยสำคัญ 0.05 เมื่อพิจารณาค่า P-value เท่ากับ 0.4119 พบว่าข้อมูลชุดนี้มีแนวโน้มของข้อมูล แสดงดังรูปภาพที่ 4-24 ดังนี้



รูปภาพ 4-24 กราฟแสดง Time series ของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของสถานีการเคหะดินแดง เขตดินแดง

4.2.2.2 ประมาณค่าพารามิเตอร์ที่เหมาะสม

หาตัวแบบจำลองที่เหมาะสมสำหรับข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของเขตดินแดงด้วยการแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) โดยเลือกรูปแบบที่เหมาะสมที่สุดสำหรับข้อมูลชุดนี้โดยใช้เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike's information criterion: AIC) และเกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criteria: BIC) แสดงดังตารางที่ 4-17

ตารางที่ 4-17 การประมาณค่าพารามิเตอร์ที่เหมาะสมของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ของสถานีการเคหะดินแดง เขตดินแดง

ค่าพารามิเตอร์	AIC	BIC
$\hat{\mu}$ =65.9725704, $\hat{\sigma}$ =19.0423487, $\hat{\xi}$ =0.1022072	-2857.684	-2844.455

ประมาณค่าพารามิเตอร์ มีค่าเท่ากับ 0.1022(0.0530) เมื่อประมาณช่วงความเชื่อมั่นจากค่าความคลาดเคลื่อนมาตรฐานจะได้ $\hat{\xi}(\mathrm{s.e})$

$$0.1022 \pm 1.96(0.0530) = (-0.00168, 0.20608)$$

เมื่อสร้างแบบจำลองค่าขีดสุดซึ่งวิเคราะห์ข้อมูลด้วยวิธีค่าเกินเกณฑ์โดยประมาณค่าพารามิเตอร์ด้วยวิธี MLE พบว่า ช่วงความเชื่อมั่นอยู่ระหว่าง -0.00168 ถึง 0.20608 จะเห็นได้ว่าช่วงความเชื่อมั่นของ $\hat{\xi}$ ครอบคลุมศูนย์ $(\hat{\xi} < 0)$ ดังนั้น ตัวแบบจำลองนี้มีการแจกแจงกัมเบล

4.2.2.3 ตรวจสอบแบบจำลอง

การตรวจสอบความเหมาะสมการแจกแจงของตัวแบบจำลองโดยใช้สถิติทดสอบ The Komolgorov-simrnov test (KS-Test) ของสถานีการเคหะดินแดง เขตดินแดง มีการแจกแจงแบบกัมเบลหรือไม่

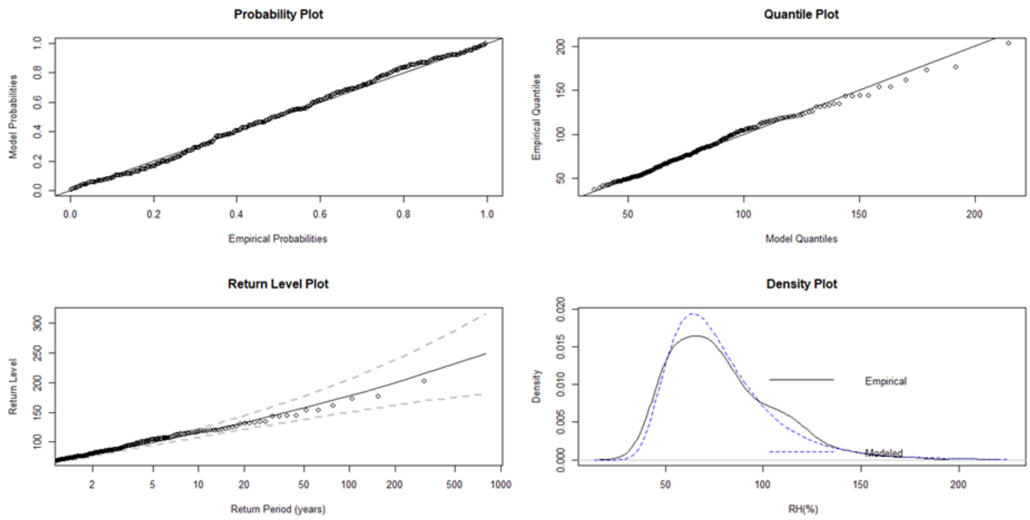
H_0 : มีการแจกแจงกัมเบล

H_1 : ไม่มีการแจกแจงกัมเบล

ระดับนัยสำคัญที่ 0.05

เนื่องจาก ค่า P-value เท่ากับ 0.7429 มากกว่า ระนัยสำคัญ 0.05 จึงทำการยอมรับสมมติฐานหลัก ดังนั้น สถานีการเคหะดินแดง เขตดินแดง มีการแจกแจงกัมเบล เป็นการแจกแจงที่เหมาะสมของตัวแบบจำลอง

นอกจากนี้ยังสามารถใช้กราฟความน่าจะเป็น (Probability Plot) และกราฟควอนไทล์ (Quantile Plot) เพื่อตรวจสอบความเหมาะสมของตัวแบบจำลองดังที่ได้แสดงในภาพที่ 4-25



รูปภาพ 4-25 กราฟแสดงฟังก์ชันการแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) ของสถานีการเคหะดินแดง เขตดินแดง

4.2.2.4 การประมาณระดับการเกิดซ้ำ

จากตัวแบบจำลองที่ได้สามารถประมาณระดับการเกิดซ้ำในช่วงเวลา ที่เกิดซ้ำใน 2, 5, 10, 50 และ 100 ปี ข้างหน้า (หน่วย ไมโครกรัม/ลูกบาศก์เมตร : วัน) แสดงได้ดังตารางที่ 4-18

ตารางที่ 4-18 แสดงค่าประมาณระดับการเกิดซ้ำ ณ ช่วงเวลาที่เกิดซ้ำต่าง ๆ

2-year level	5-year level	10-year level	50-year level	100-year level
73.08116	96.83936	114.15474	157.28457	177.82600

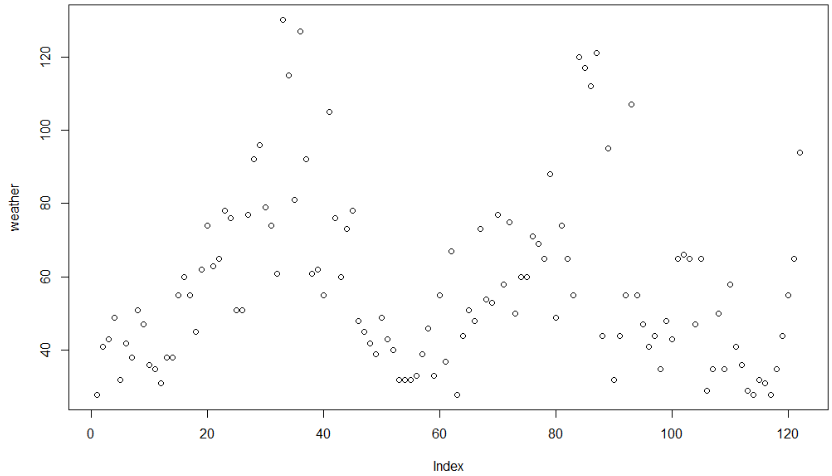
(ไมโครกรัม/ลูกบาศก์เมตร (μg/m3))

จากตารางที่ 4-18 พบว่า เมื่อรอบปีการเกิดซ้ำเพิ่มขึ้นจะทำให้ค่าระดับการเกิดซ้ำเพิ่มขึ้นตามไปด้วย โดยตัวแบบจำลองนี้สามารถใช้คาดการณ์ระดับการเกิดซ้ำของปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของสถานีการเคหะดินแดง เขตดินแดง ในรอบปีที่ต้องการทราบได้ ทำให้สามารถนำไปปรับใช้วางแผนควบคุมปริมาณความเข้มข้นของ $PM_{2.5}$ เพื่อคุณภาพชีวิตที่ดีของประชาชน

4.2.3 สถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน

4.2.3.1 การเตรียมข้อมูล

1. นำข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของสถานีการเคหะดินแดง เขตดินแดง แสดงผลดังรูปภาพที่ 4-26



รูปภาพ 4-26 กราฟแสดงการกระจายของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน

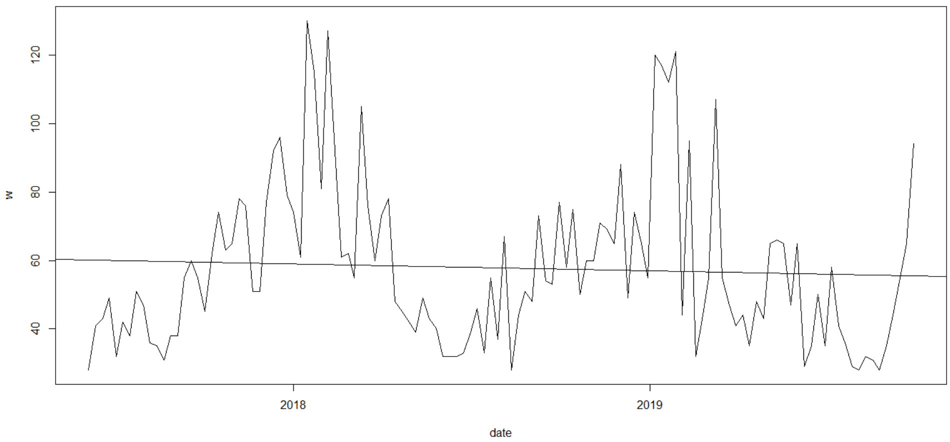
2. การทดสอบความคงที่ (Stationary) และแนวโน้ม (Trend) ของข้อมูล

ตารางที่ 4-19 การทดสอบความคงที่และแนวโน้มของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ของสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน

การทดสอบ	ค่าสถิติ	P-value
Augmented Dickey-Fuller (ADF) test	-2.7743	0.2549
Mann-Kendall trend test	-0.83033	0.4064

*(*ระดับนัยสำคัญ 0.05)*

จากตารางที่ 4-19 แสดงการทดสอบ จากวิธีการทดสอบของ Augmented Dickey-Fuller (ADF) test ที่ระดับนัยสำคัญ 0.05 เมื่อพิจารณาค่า P-value เท่ากับ 0.2549 พบว่าข้อมูลชุดนี้ไม่อยู่ภายใต้กระบวนการคงที่ และพิจารณาจากวิธีทดสอบของ Mann-Kendall trend (MK) test ที่ระดับนัยสำคัญ 0.05 เมื่อพิจารณาค่า P-value เท่ากับ 0.4064 พบว่าข้อมูลชุดนี้มีแนวโน้มของข้อมูล แสดงดังภาพที่ 4-27 ดังนี้



รูปภาพ 4-27 กราฟแสดง Time series ของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ ของสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน

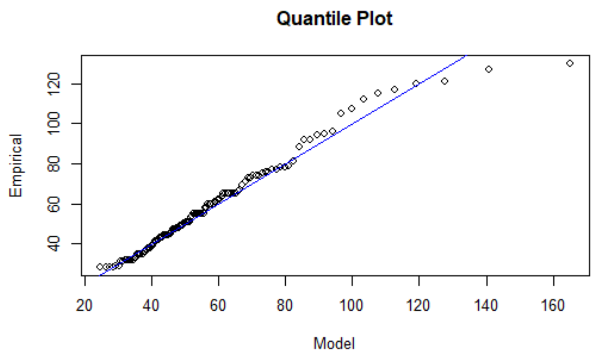
4.2.3.2 ประมาณค่าพารามิเตอร์ที่เหมาะสม

การหาตัวแบบจำลองที่เหมาะสมที่สุดสำหรับข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ ทำได้โดยการหาค่าประมาณพารามิเตอร์ ช่วงความเชื่อมั่น และการแจกแจงของตัวแบบจำลอง จากวิธีการแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) แล้วคัดเลือกตัวแบบที่เหมาะสมที่สุดโดยใช้เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike's information criterion: AIC) และเกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criteria: BIC) แสดงดังตารางที่ 4-20

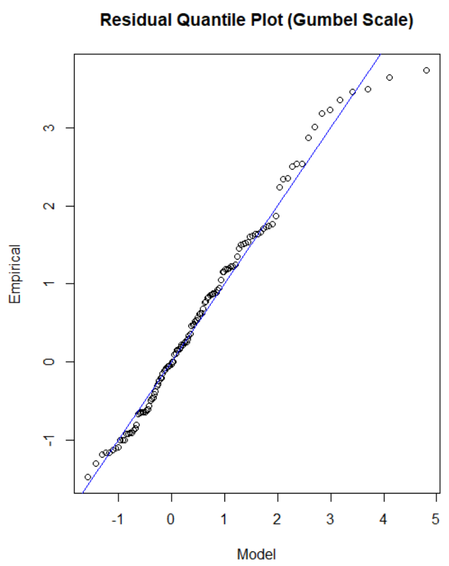
ตารางที่ 4-20 การประมาณค่าพารามิเตอร์ที่เหมาะสมของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ของสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน

รูปแบบ	ค่าพารามิเตอร์	AIC	BIC
1	μ =41.444, $\hat{\sigma}$ =14.662, $\hat{\xi}$ =0.220	-1073.835	-1065.423
2	$\mu(t)$ =47.73860672 -0.03215919t, $\hat{\sigma}$ =15.38613693, $\hat{\xi}$ =0.18859061	-1070.698	-1059.482
3	$\mu(t)$ = 47.9548105670 - 0.0004287688t - 0.0004157332t^2 , σ = 15.2712546887 , $\hat{\xi}$ = 0.1880522342	-	-
4	-	-	-
5	-	-	-

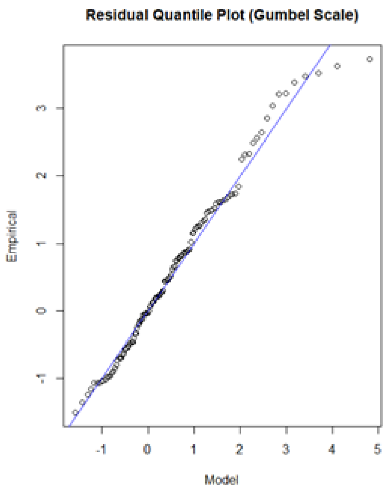
จากภาพที่ 4-21 ถ้าตัวแบบจำลองมีการแจกแจงที่เหมาะสม Q-Q plots จะอยู่ใกล้เส้นทแยงมุมซึ่งกราฟจะเป็นเส้นตรง พบว่า ตัวแบบจำลองที่ 1-2 มีการแจกแจงที่เหมาะสม เนื่องจาก Q-Q plots เกาะกลุ่มอยู่ใกล้กับเส้นทแยงมุมและกราฟเป็นเส้นตรง และตัวแบบจำลองที่ 3 ถึงแม้ว่าจะมี Q-Q plots เกาะกลุ่มอยู่ใกล้กับเส้นทแยงมุมและกราฟเป็นเส้นตรง แต่มีการแจกแจงที่ไม่เหมาะสมเนื่องจากคำนวณค่า Standard Error (S.E) ไม่ได้ หมายความว่าค่าประมาณพารามิเตอร์ของตัวแบบจำลองที่ 3 เมื่อนำไปพยากรณ์แล้วมีค่าผิดพลาดเยอะ ทำให้นำไปพยากรณ์ได้ไม่แม่นยำ ส่วนตัวแบบจำลองที่ 4-5 มีการแจกแจงที่ไม่เหมาะสม เนื่องจาก Q-Q plots ไม่อยู่ใกล้กับเส้นทแยงมุมและกราฟไม่เป็นเส้นตรง ดังนั้นตัวแบบจำลองที่จะนำไปพิจารณาหาการแจกแจงที่เหมาะสมที่สุด คือ ตัวแบบจำลองที่ 1-2 และจากตารางที่ 4-20 เมื่อพิจารณาค่า AIC และ BIC พบว่า ตัวแบบที่ 1 เป็นแบบจำลองที่ให้ ค่า AIC และ BIC น้อยที่สุด นั้นคือ เป็นตัวแบบที่เหมาะสมที่สุดข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ของสถานีสถานีตำรวจนครบาลโชคชัยกรุงเทพมหานคร เขตวังทองหลาง ดังนั้น สามารถเขียนอธิบายได้ว่า $x \sim GEV(\mu,\sigma,\xi)$ โดยที่พารามิเตอร์ μ =41.444, $\hat{\sigma}$ =14.662 และ $\hat{\xi}$ =0.220 มีค่าคงที่



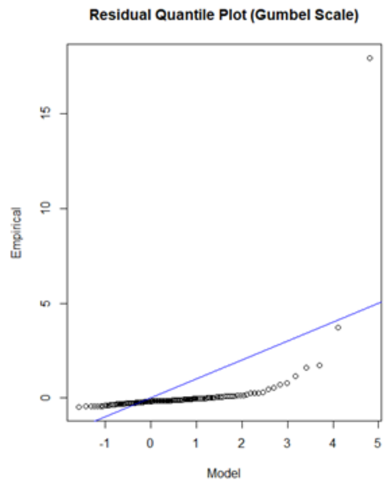
ตัวแบบที่ 1



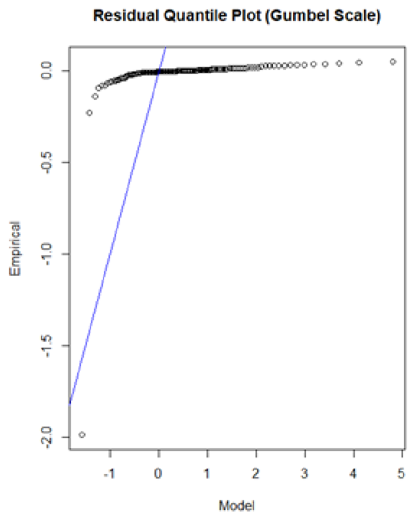
ตัวแบบที่ 2



ตัวแบบที่ 3



ตัวแบบที่ 4



ตัวแบบที่ 5

รูปภาพ 4-28 กราฟ Q-Q Plot ของตัวแบบ

ประมาณค่าพารามิเตอร์ มีค่าเท่ากับ 0.220(0.0951) เมื่อประมาณช่วงความเชื่อมั่นจากค่าความคลาดเคลื่อนมาตรฐานจะได้ $\hat{\xi}(\mathrm{s.e})$

$$0.220 \pm 1.96(0.0951) = (0.0336, 0.4064)$$

เมื่อสร้างแบบจำลองค่าขีดสุดซึ่งวิเคราะห์ข้อมูลด้วยวิธีค่าเกินเกณฑ์โดยการประมาณค่าพารามิเตอร์ด้วยวิธี MLE พบว่า ช่วงความเชื่อมั่นอยู่ระหว่าง 0.033604 ถึง0.406396 จะเห็นได้ว่าช่วงความเชื่อมั่นของ $\hat{\xi}$ มากกว่าศูนย์ $(\hat{\xi} > 0)$ ดังนั้น ตัวแบบจำลองนี้มีการแจกแจงฟรีเชท

4.2.3.3 ตรวจสอบแบบจำลอง

การตรวจสอบความเหมาะสมการแจกแจงของตัวแบบจำลองโดยใช้สถิติทดสอบ The Komolgorov-simrnov test (KS-Test) ของสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน มีการแจกแจงฟรีเชท หรือไม่

สมมติฐานทดสอบ

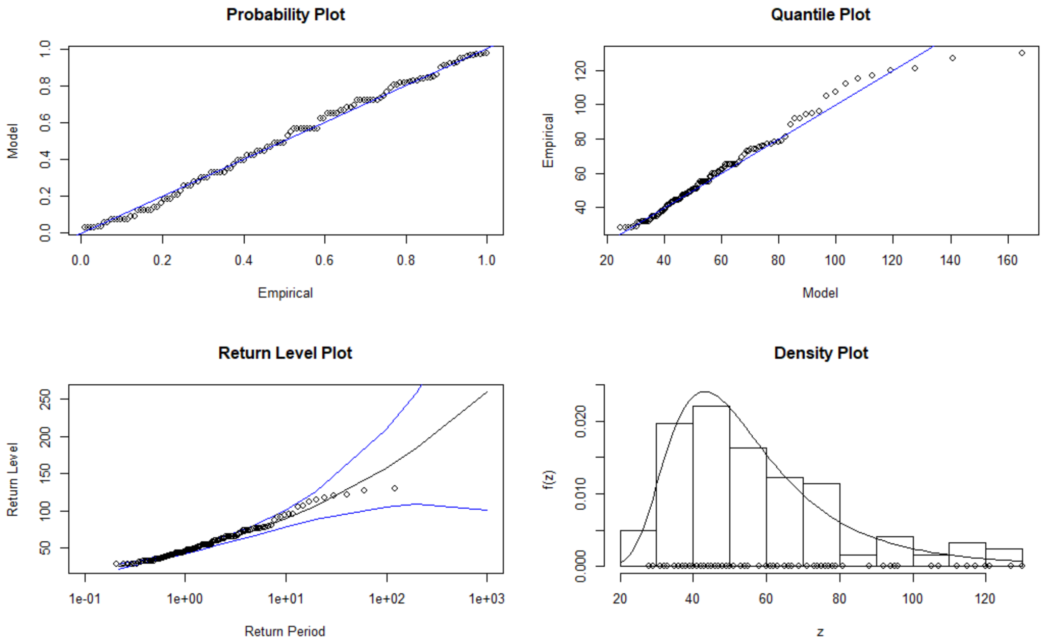
H_0 : มีการแจกแจงฟรีเชท

H_1 : ไม่มีการแจกแจงฟรีเชท

ระดับนัยสำคัญที่ 0.05

เนื่องจาก ค่า P-value เท่ากับ 0.8798 มากกว่า ระนัยสำคัญ 0.05 จึงทำการยอมรับสมมติฐานหลัก ดังนั้น สถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน มีการแจกแจงฟรีเชท เป็นการแจกแจงที่เหมาะสมของตัวแบบจำลอง

นอกจากนี้ยังสามารถใช้กราฟความน่าจะเป็น (Probability Plot) และกราฟควอนไทล์ (Quantile Plot) เพื่อตรวจสอบความเหมาะสมของตัวแบบจำลองดังที่ได้แสดงในรูปภาพที่ 4-29



รูปภาพ 4-29 กราฟแสดงฟังก์ชันการแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) ของสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน

4.2.3.4 การประมาณระดับการเกิดซ้ำ

จากตัวแบบจำลองที่ได้ สามารถประมาณระดับการเกิดซ้ำของปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของเขตปทุมวัน ในอีก 2, 5, 10, 50 และ 100 ปี ข้างหน้า (หน่วย ไมโครกรัม/ลูกบาศก์เมตร : วัน) ได้ดังตารางที่ 4-21

ตารางที่ 4-21 แสดงค่าประมาณระดับการเกิดซ้ำ ณ ช่วงเวลาที่เกิดซ้ำต่าง ๆ

Station	Model	2-year level	5-year level	10-year level	50-year level	100-year level
Patumwan	1	108.3851	116.6103	121.4974	129.6083	132.0663

(ไมโครกรัม/ลูกบาศก์เมตร (μg/m3))

จากตารางที่ 4-21 เมื่อรอบปีการเกิดซ้ำเพิ่มขึ้นจะทำให้ค่าระดับการเกิดซ้ำเพิ่มขึ้นตามไปด้วย โดยตัวแบบจำลองนี้สามารถใช้คาดการณ์ระดับการเกิดซ้ำของปริมาณความเข้มข้นของ $PM_{2.5}$ รายสัปดาห์ของสถานีโรงพยาบาลจุฬาลงกรณ์ เขตปทุมวัน ในรอบปีที่ต้องการทราบได้ ทำให้สามารถนำไปปรับใช้วางแผนควบคุมปริมาณความเข้มข้นของ $PM_{2.5}$ เพื่อคุณภาพชีวิตที่ดีของประชาชน

บทที่ 5
สรุปผลการดำเนินงานและข้อเสนอแนะ

การดำเนินการศึกษาครั้งนี้มีวัตถุประสงค์เพื่อหาตัวแบบจำลองที่เหมาะสมและระดับการเกิดซ้ำของปริมาณความเข้มข้นของ $PM_{2.5}$ ใน 3 เขตธุรกิจของกรุงเทพมหานคร ได้แก่ เขตวังทองหลาง เขตดินแดง และเขตปทุมวัน ด้วยวิธีการแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution : GEV) และวิธีการแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution : GPD) และเปรียบเทียบตัวแบบจำลองของทั้งสองวิธี โดยใช้เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike's information criterion : AIC) และเกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criteria : BIC)

5.1 สรุปผลการดำเนินงาน

ผลจากการนำค่าประมาณพารามิเตอร์แต่ละเขตมาหาช่วงความเชื่อมั่นการแจกแจงเพื่อหาการแจกแจงของตัวแบบจำลองที่เหมาะสมของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ รายชั่วโมงและรายสัปดาห์ ใน 3 เขตธุรกิจของกรุงเทพมหานคร ได้แก่ เขตวังทองหลาง เขตดินแดง และเขตปทุมวัน โดยวิธีการแจกแจงพาเรโตวางนัยทั่วไป (GPD) พบว่า

เขตวังทองหลาง มีตัวแบบจำลองภายใต้การแจกแจงพาเรโต II
เขตดินแดง มีตัวแบบจำลองภายใต้การแจกแจงเลขชี้กำลัง
เขตปทุมวัน มีตัวแบบจำลองภายใต้การแจกแจงพาเรโต II

แสดงดังตารางที่ 5-1

และวิธีการแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) พบว่า

เขตวังทองหลาง มีตัวแบบจำลองภายใต้การแจกแจงฟรีเชท
เขตดินแดง มีตัวแบบจำลองภายใต้การแจกแจงกัมเบล
เขตปทุมวัน มีตัวแบบจำลองภายใต้การแจกแจงฟรีเชท

แสดงดังตารางที่ 5-2

ตารางที่ 5-1 แสดงช่วงความเชื่อมั่นของ ξ ภายใต้การแจกแจงพาเรโตวางนัยทั่วไป (GPD) ของทั้ง 3 เขตในกรุงเทพมหานคร ที่ระดับนัยสำคัญ 0.05

เขต	ขอบล่าง	ขอบบน	P-value	การแจกแจง
เขตวังทองหลาง	-0.06554	-0.29202	0.2527	พาเรโต II
เขตดินแดง	-0.18376	0.11599	0.832	เลขชี้กำลัง
เขตปทุมวัน	-0.43337	-0.16331	0.571	พาเรโต II

ตารางที่ 5-2 แสดงช่วงความเชื่อมั่นของ ξ ภายใต้การแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) ของทั้ง 3 เขตในกรุงเทพมหานคร ที่ระดับนัยสำคัญ 0.05

เขต	ขอบล่าง	ขอบบน	P-value	การแจกแจง
เขตวังทองหลาง	0.0656	0.3745	0.9071	ฟรีเซท
เขตดินแดง	-0.00168	0.20608	0.7429	กัมเบล
เขตปทุมวัน	0.0336	0.4064	0.8798	ฟรีเซท

ตารางที่ 5-3 การเปรียบเทียบของตัวแบบของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ใน 3 เขตธุรกิจของกรุงเทพมหานครภายใต้การแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) และการแจกแจงพาเรโตวางนัยทั่วไป (GPD) ด้วยเกณฑ์ AIC และ BIC

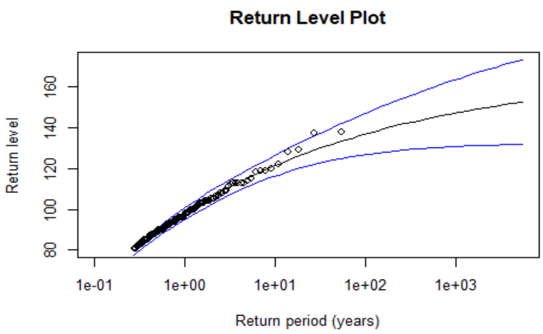
เขต	GPD		GEV	
	AIC	BIC	AIC	BIC
เขตวังทองหลาง	**-1416.525**	**-1409.938**	-1065.656	-1057.244
เขตดินแดง	-1401.122	-1394.638	**-2857.684**	**-2844.455**
เขตปทุมวัน	**-1413.355**	**-1406.798**	-1073.835	-1065.423

จากตารางที่ 5-3 เป็นการเปรียบเทียบตัวแบบจำลองที่เหมาะสมที่สุดของของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ใน 3 เขตธุรกิจของกรุงเทพมหานคร ได้แก่ เขตดินแดง เขตปทุมวัน และเขตวังทองหลาง ระหว่างวิธีการแจกแจงพาเรโตวางนัยทั่วไป (GPD) และวิธีการแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) โดยใช้เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike’s information criterion : AIC) และเกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criteria : BIC) พบว่า วิธีการแจกแจงพาเรโตวางนัยทั่วไป (GPD) เป็นวิธีที่เหมาะสมที่สุดเพื่อใช้หาตัวแบบจำลองของเขตวังทองหลาง และเขตปทุมวัน เนื่องจากค่า AIC และ BIC ของสองวิธีนี้น้อยที่สุดและวิธีการแจกแจงค่าสุดขีดวางนัยทั่วไป (GEV) เป็นวิธีที่เหมาะสมที่สุดเพื่อใช้หาตัวแบบจำลองของเขตดินแดง เนื่องจากค่า AIC และ BIC ของวิธีนี้น้อยที่สุด ดังนั้น เขตวังทองหลาง และเขตปทุมวันจึงมีตัวแบบจำลองที่เหมาะสมที่สุดภายใต้การแจกแจงพาเรโต II ส่วนเขตดินแดงมีตัวแบบจำลองที่เหมาะสมที่สุดภายใต้การแจกแจงกัมเบล

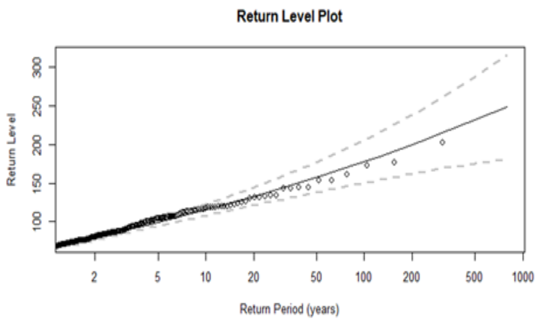
ตารางที่ 5- 4 แสดงช่วงระดับการเกิดซ้ำในรอบ 2 ปี 5 ปี 10 ปี 50 ปี และ 100 ปี ข้างหน้า

เขต	2-year level	5-year level	10-year level	50-year level	100-year level
เขตวังทองหลาง	105.9907	115.2559	121.3220	132.8173	136.8352
เขตดินแดง	73.08116	96.83936	114.15474	157.28457	177.82600
เขตปทุมวัน	108.3851	116.6103	121.4974	129.6083	132.0663

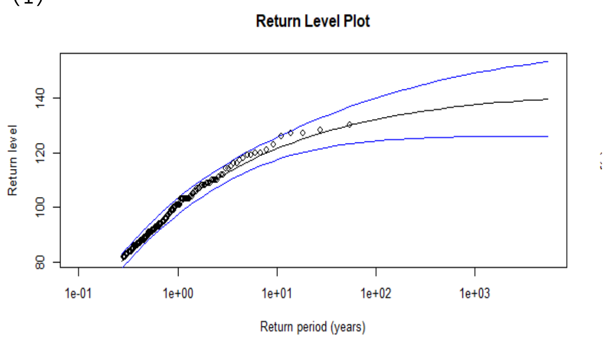
จากตารางที่ 5-4 พิจารณาระดับการเกิดซ้ำในรอบ 2 ปี 5 ปี 10 ปี 50 ปี และ 100 ปี ข้างหน้า ของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ในทั้ง 3 เขตธุรกิจของกรุงเทพมหานคร ได้แก่ เขตดินแดง เขตปทุมวัน และเขตวังทองหลาง พบว่า ทั้ง 3 เขต เมื่อมีรอบปีการเกิดซ้ำเพิ่มมากขึ้น จะทำให้ค่าระดับการเกิดซ้ำเพิ่มขึ้นตามไปด้วย



(1)



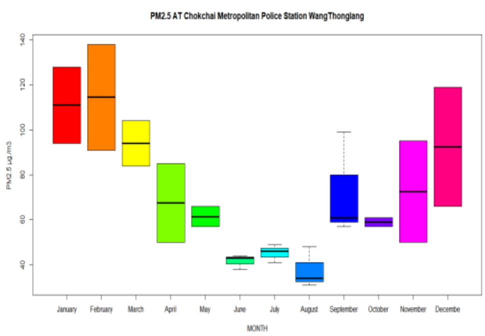
(2)



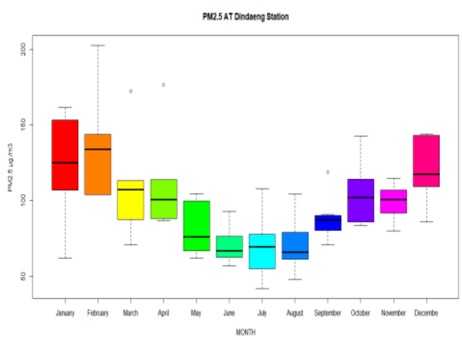
(3)

รูปภาพ 5-1 กราฟแสดง Return Level Plot (1) เขตวังทองหลาง (2) เขตดินแดง และ (3) เขตปทุมวัน

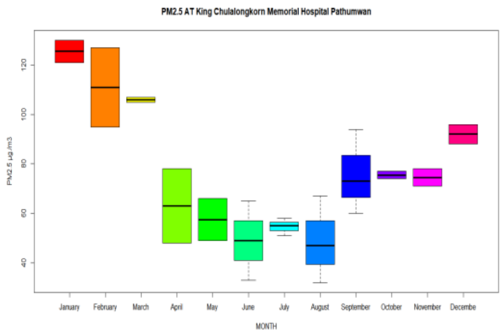
จากภาพที่ 5-1 พบว่า สามารถวิเคราะห์ระดับการเกิดซ้ำของเขตวังทองหลางและเขตปทุมวันได้ดีในช่วงรอบปีไม่เกิน 10 ปี และมีปริมาณความเข้มข้นของ $PM_{2.5}$ ไม่เกิน 120 ไมโครกรัม/ลูกบาศก์เมตร ส่วนเขตดินแดงมีระดับการเกิดซ้ำได้ดีในช่วงรอบปีไม่เกิน 20 ปี และมีปริมาณความเข้มข้นของ $PM_{2.5}$ ไม่เกิน 140 ไมโครกรัม/ลูกบาศก์เมตร โดยที่เขตดินแดงนั้นมีระดับการเกิดซ้ำสูงกว่าเขตวังทองหลางและเขตปทุมวัน โดยเกณฑ์ที่ใช้วัดปริมาณความเข้มข้นของ $PM_{2.5}$ ของกรมควบคุมมลพิษได้กำหนดค่ามาตรฐานเฉลี่ยไว้ไม่เกิน 50 ไมโครกรัม/ลูกบาศก์เมตร หมายความว่าปริมาณความเข้มข้นของ $PM_{2.5}$ ตั้งแต่ 50 ไมโครกรัม/ลูกบาศก์เมตรขึ้นไปจะเริ่มมีผลต่อสุขภาพของประชาชน โดยทั้ง 3 เขตมีปริมาณความเข้มข้นของ $PM_{2.5}$ อยู่ในระดับเกินเกณฑ์มาตรฐานทุกเขต และฝุ่นละออง $PM_{2.5}$ ก็มีแนวโน้มเพิ่มสูงขึ้นในอนาคต เนื่องจากมีสาเหตุมาจากการคมนาคม และการจราจรที่หนาแน่น การสร้างตึกและอาคาร รวมถึงที่อยู่อาศัยที่อยู่กันอย่างอย่างแออัด การสะสมฝุ่นควันพิษจากการเผาไหม้จากยานพาหนะ และเชื้อเพลิงในครัวเรือนและวัชพืชจากทั้งในกรุงเทพมหานครและพื้นที่ใกล้เคียง อาทิเช่น การเผาป่าของจังหวัดใกล้เคียง การเผาไหม้ทางการเกษตรของประเทศเพื่อนบ้าน เป็นต้น รวมทั้งสภาพอากาศที่มีสภาวการณ์ผันกลับของอุณหภูมิ จากการนำข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ มาวิเคราะห์ปริมาณความเข้มข้นของ $PM_{2.5}$ มักจะเข้มข้นขึ้นในช่วงฤดูหนาว (พฤศจิกายน - มีนาคม) ดังภาพที่ 5-2 ซึ่งมีสาเหตุมาจากในช่วงฤดูหนาวนั้น อากาศชั้นความเย็นจะถูกกักอยู่ภายใต้อากาศชั้นความร้อนส่งผลให้ฝุ่นละอองถูกกักกั้นทำให้ไม่สามารถลอยตัวสูงได้ จึงรวมตัวกันอยู่ในชั้นอากาศที่มนุษย์เราอาศัยอยู่ ดังนั้น เพื่อควบคุมคุณภาพอากาศให้อยู่ในระดับที่พอดีและไม่ส่งผลกระทบต่อสุขภาพของประชาชน จะต้องดำเนินการแก้ไขปัญหาให้ตรงจุด ถึงแม้ว่าปริมาณความเข้มข้นของ $PM_{2.5}$ จะเกินค่ามาตรฐานมาเพียงเล็กน้อยก็ส่งผลกระทบต่อสุขภาพอนามัยของประชาชน และอาจนำไปสู่การปรับค่ามาตรฐานให้เข้มงวดขึ้นเพื่อควบคุมคุณภาพอากาศ ทั้งหมดนี้ไม่ได้หมายความถึงเพียงแค่ภายใน 3 เขตของกรุงเทพมหานครนี้เท่านั้น แต่รวมไปถึงเขตอื่นๆที่มีประชากรดำเนินชีวิตอยู่ด้วย



(1)



(2)



(3)

รูปภาพ 5-2 แสดง Box Plot ปริมาณความเข้มข้นของ $PM_{2.5}$ สูงสุดรายเดือน (1) เขตวังทองหลาง (2) เขตดินแดง และ (3) เขตปทุมวัน

5.2 ข้อเสนอแนะ

ผู้ศึกษาโครงงานพิเศษได้เสนอแนวทางแก้ไขปัญหาดังกล่าวเพื่อประโยชน์ในการศึกษาในอนาคต ดังนี้

5.2.1 เมื่อสามารถเลือกตัวแบบจำลองที่มีความเหมาะสมกับลักษณะของข้อมูลปริมาณความเข้มข้นของ $PM_{2.5}$ ของ 3 เขตในกรุงเทพมหานครได้นั้นจะสามารถคาดการณ์ระดับการเกิดซ้ำในรอบ 2, 5, 10, 50 และ 100 ปี ข้างหน้า ซึ่งสามารถนำไปใช้เป็นข้อมูลเพิ่มเติมในการเตรียมวางแผนรับมือและจัดการกับปัญหาฝุ่นละออง $PM_{2.5}$ ในกรุงเทพมหานครเมื่อเกินค่ามาตาราฐานที่กำหนดได้

5.2.2 ข้อมูลที่นำมาศึกษามีเพียงปริมาณความเข้มข้นของ $PM_{2.5}$ โดยไม่ได้ศึกษาถึงปัจจัยอื่นๆ ที่อาจมีผลต่อปริมาณความเข้มข้นของ $PM_{2.5}$ ดังนั้น การเลือกข้อมูลมาวิเคราะห์จึงควรมีความเหมาะสมกับการแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution: GEV) และการแจกแจงพาเรโตวางนัยทั่วไป (Generalized Pareto Distribution: GPD)

บรรณานุกรม

ปิยภัทร บุษบาบดินทร์, และ อรุณ แก้วมั่น. (2015). สถิติค่าสุดขีด. *วารสารวิชาการพระจอมเกล้าพระนครเหนือ, 2(25)*, 315-324.

ปิยภัทร บุษบาบดินทร์. (2017). การวิเคราะห์ค่าสุดขีดด้วย R. มหาสารคาม: โรงพิมพ์มหาวิทยาลัยราชภัฏมหาสารคาม.

วิชุดำ เห็นเจริญ, และ ปิยภัทร บุษบาบดินทร์. (2017). แบบจำลองค่ำสุดขีดของอัตราการป่วยโรคไข้หวัดใหญ่ในภาคตะวันออกเฉียงเหนือของประเทศไทย. *วารสารวิทยาศาสตร์บูรพา, 3(22)*, 123-140

อรุณ แก้วมัน, เสาวนีย์ รัตนะวัน, และ ปิยภัทร บุษบาบดินทร์. (2014). การสร้างแบบจำลองค่าสุดขีดปริมาณฝนสูงสุดในภาคตะวันออกเฉียงเหนือตอนล่างของประเทศไทย. *วารวิทยาศาสตร์ประยุกต์, 2(13)*, 55-65.

Guanghui, Y., Feifan, W., Jing, H.,Yan, L., & Xianzhao, L. (2019). Value Assessment of Health Losses Caused by PM2.5 in Changsha City, China.International ournal of Environmental Research and Public Health. The School of Resource, Environment and Safety Engineering, Hunan University of Science and Technology. China

Martins, L.D., Wikuats, C.F., Capucim, M.N. et al. (2017). Extreme value analysis of air pollution data and their comparison between two large urban of South America. Journal of Elsevier: Weather and Climate Extremes. Brazil

ภาคผนวก ก

ตัวแบบ GPD รายชั่วโมง

CHOKCHAI Metropolitan Police Station WangThonglang PM2.5

การเรียกใช้แพคเกจของโปรแกรม

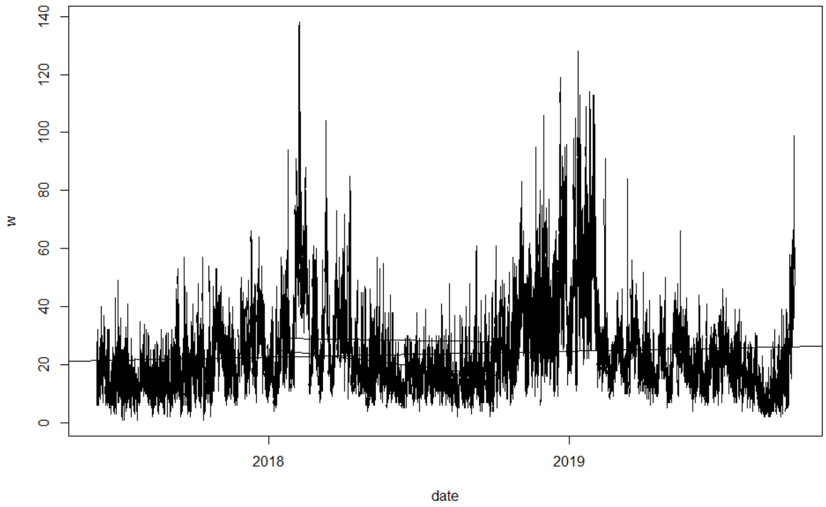
```
library(extRemes)
library(ismev)
library(aTSA)
library(tseries)
library(trend)
library(e1071)
library(TLMoments)
library(nortest)
library(eva)
library(ADGofTest)
library(lattice)
```

การเรียกใช้ข้อมูล

```
data <-
read.csv("C:\\Users\\acer\\Desktop\\Project\\Code\\clean_ChokchaiMetropolitan
PoliceStation_WangThonglang\\clean_ChokchaiMetropolitanPoliceStation_WangThong
lang_pm2.5_hour.csv", header = T)
```

การสร้างกราฟเพื่อดูแนวโน้ม

```
date <- as.POSIXct(data$DAY, format = "%d/%m/%Y %H:%M")
linearmodel <- lm(data$PM2.5 ~ date)
plot(data$PM2.5 ~ date, type = 'l')
abline(linearmodel)
weather = na.omit(data$PM2.5)
```



การทดสอบเพื่อดูแนวโน้มของข้อมูล

mk.test(weather, continuity = T)

Mann-Kendall trend test

data: weather

z = 13.798, n = 20146, p-value < 2.2e-16

alternative hypothesis: true S is not equal to 0

sample estimates:

S varS tau

1.314605e+07 9.077641e+11 6.566212e-02

หา summary

summary(weather)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	13.00	20.00	23.76	29.00	138.00

การทดสอบความเป็น Stationary ความเบ้ ความโค้ง

stationary.test(weather)

Augmented Dickey-Fuller Test

alternative: stationary

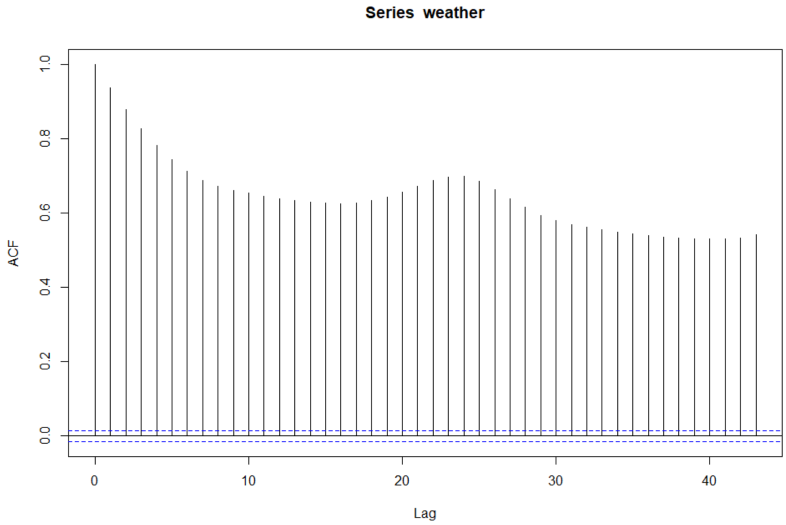
Type 1: no drift no trend

```
      lag    ADF p.value
 [1,]   0 -13.51    0.01
 [2,]   1 -13.26    0.01
 [3,]   2 -12.61    0.01
 [4,]   3 -11.84    0.01
 [5,]   4 -11.22    0.01
 [6,]   5 -10.55    0.01
 [7,]   6  -9.75    0.01
 [8,]   7  -8.92    0.01
 [9,]   8  -8.38    0.01
[10,]   9  -7.86    0.01
[11,]  10  -7.61    0.01
[12,]  11  -7.26    0.01
[13,]  12  -6.81    0.01
[14,]  13  -6.48    0.01
```

Type 2: with drift no trend

```
      lag   ADF p.value
 [1,]   0 -25.3    0.01
 [2,]   1 -25.1    0.01
 [3,]   2 -24.1    0.01
 [4,]   3 -22.9    0.01
 [5,]   4 -21.8    0.01
 [6,]   5 -20.7    0.01
 [7,]   6 -19.3    0.01
 [8,]   7 -17.8    0.01
 [9,]   8 -16.8    0.01
[10,]   9 -15.9    0.01
[11,]  10 -15.4    0.01
```

```
[12,]  11 -14.8    0.01
[13,]  12 -14.0    0.01
[14,]  13 -13.4    0.01
Type 3: with drift and trend
      lag    ADF p.value
 [1,]   0 -25.4    0.01
 [2,]   1 -25.2    0.01
 [3,]   2 -24.2    0.01
 [4,]   3 -23.0    0.01
 [5,]   4 -21.9    0.01
 [6,]   5 -20.8    0.01
 [7,]   6 -19.4    0.01
 [8,]   7 -17.8    0.01
 [9,]   8 -16.9    0.01
[10,]   9 -15.9    0.01
[11,]  10 -15.5    0.01
[12,]  11 -14.9    0.01
[13,]  12 -14.0    0.01
[14,]  13 -13.4    0.01
----
Note: in fact, p.value = 0.01 means p.value <= 0.01
skewness(weather)
[1] 1.891857
kurtosis(weather)
[1] 5.0557
acf(weather)
```



adf.test(weather)

Augmented Dickey-Fuller Test

data: weather

Dickey-Fuller = -8.9339, Lag order = 27, p-value = 0.01

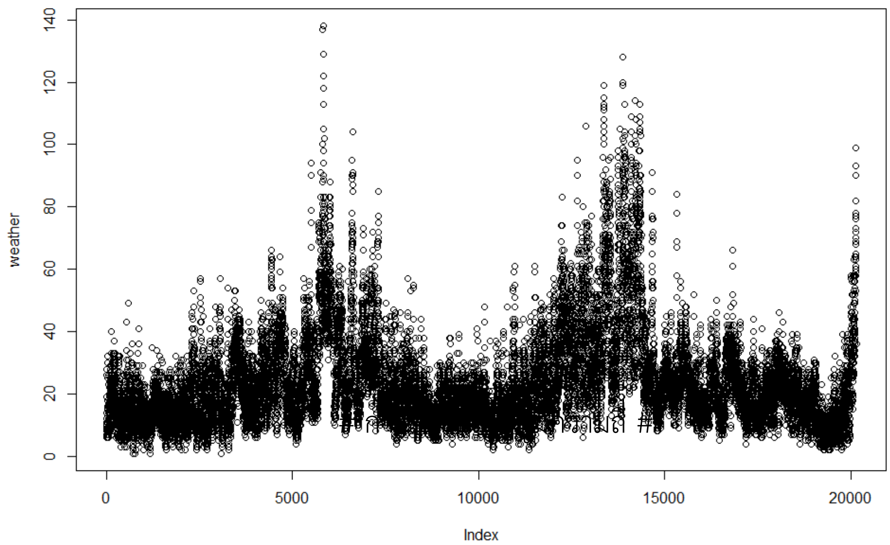
alternative hypothesis: stationary

Warning message:

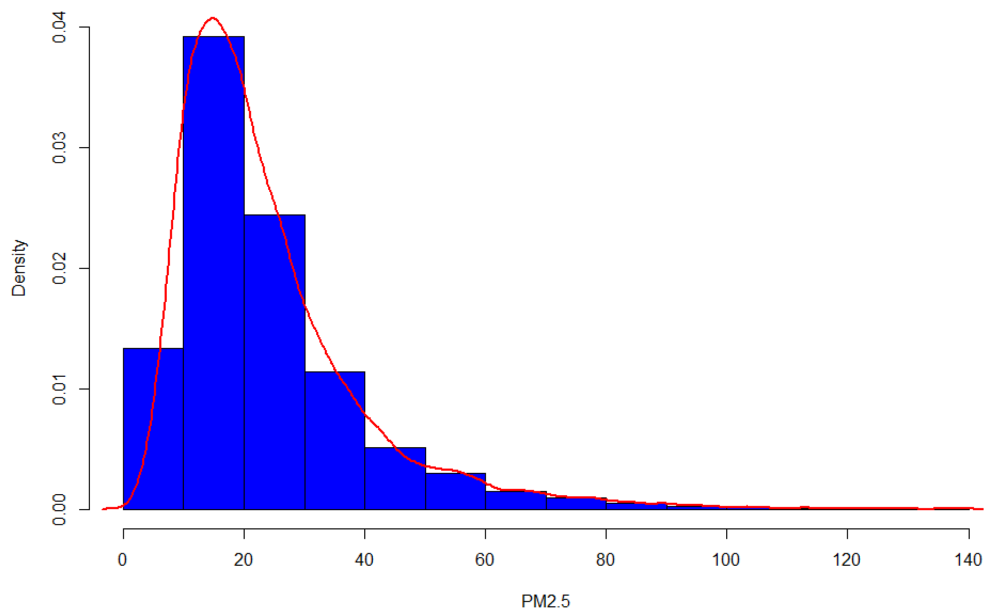
In adf.test(weather) : p-value smaller than printed p-value

การสร้างกราฟเพื่อดูการกระจายตัวของข้อมูล

plot(weather)

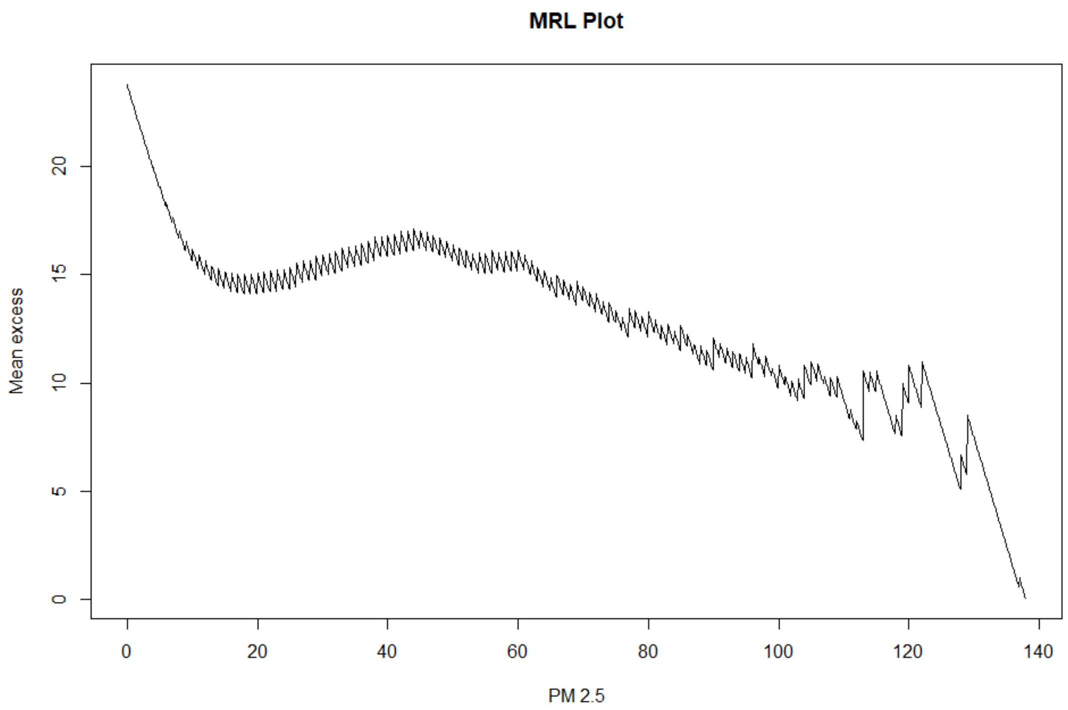


```
par(mfrow = c(1,1))
plot.ts(weather,
        xlab = "Days",
        ylab = "original data")
hist(weather,
     xlab  = "PM2.5",
     ylim  = c(0, 0.04),
     main  = " ",
     col   = "blue",
     border = "black",
     prob  = TRUE)
lines(density(weather),
      xlab  = " ",
      main = " ",
      col   = "red",
      lty    = 1,
      lwd  = 2)
```



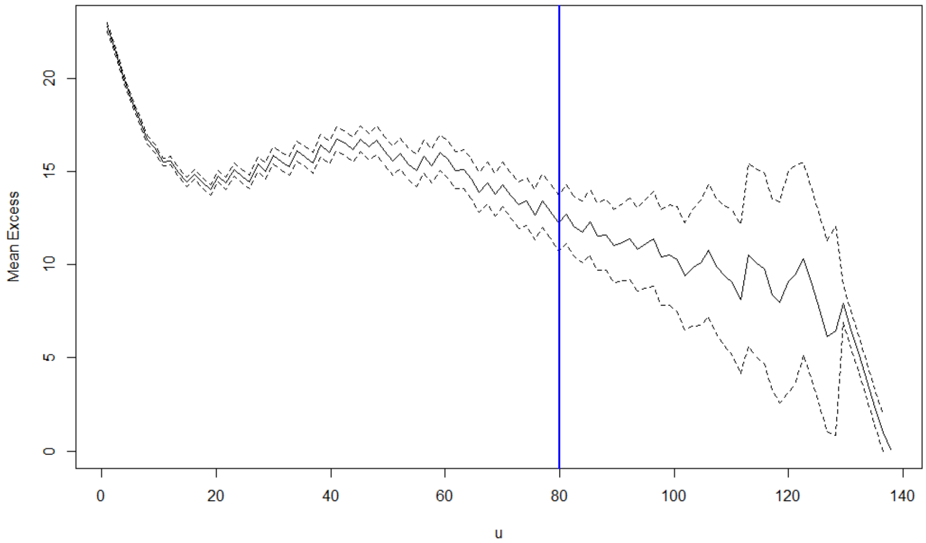
การสร้างกราฟ MRL

```
u <- seq (0, max(weather), 0.1)
x <- vector("numeric", length(u))

for (i in 1 : length (x) ) {
  threshold.exceedances <- weather[weather > u[i]]
  x[i] <- mean(threshold.exceedances - u[i])
}
plot(x ~ u,
     xlab = "PM 2.5",
     ylab = "Mean excess",
     main = "MRL Plot",
     type = "l")
```



การกำหนดค่าเกณฑ์บนกราฟ MRL

```
mrl.plot(weather)
abline(v = 80,
       col = "blue",
 lwd = 2)
```



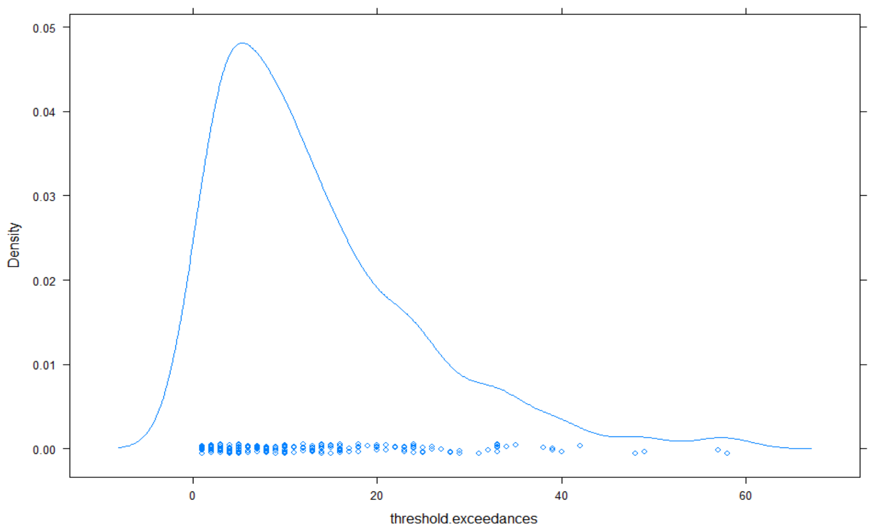
หาค่า Percentile

```
duration = data$PM2.5
quantile(duration, c( .95, .96, .97, .98, .99))
95%   96%   97%  98%  99%
 55    58    63    70   80
```

Data pre-processing

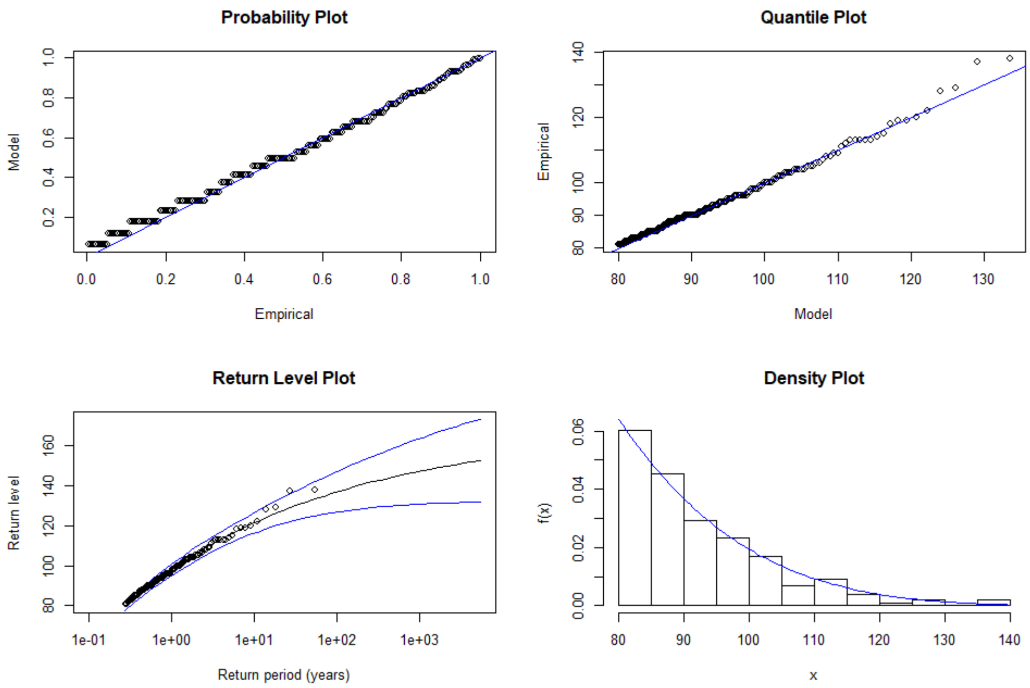
```
u <- 80
above.threshold <- weather[weather > u]
length(above.threshold)
[1] 199
threshold.exceedances <- above.threshold - u
above.threshold
```

```
  [1]  90  94  83  91  81  88 100 137 118  98  83  90  89  94
 [15]  94 105 122 129 138 113 102  81  83  83  83  81  88  81
 [29]  90  89  95  90 104  91  87  85  85  83  82  90  95 106
 [43]  96  82  82  90  88 100 106 112 108 113 115 102  82  88
 [57]  85 100 104 119 111  86  84  92  86  86  83  85  87  88
 [71]  88  83  95  84  83  82  83  84  83  85  81  82  96  89
 [85]  82  84  98  96  87  92  90  96  93  96  86  81  81  81
 [99] 105  85  91  85  90  95  99  81 102 128 120 119  92  87
[113] 101  96  88  84  83  89  85  98  97  97  85  89  86  92
[127]  96 103 104 113  91  83  94  87  85  93  96  88  85  90
[141]  95  89  89  92  93 104  96 100 109  94  82  82  92  91
[155] 103 114 101  84  83  90  87  93 108 100  94  84  86  93
[169]  89  98  85  88  98 109 113 107 104 113 105  88  82  90
[183]  83  85  86  84  81  90 103 103  85  90  85  91  84  93
[197]  99  90  82
densityplot(threshold.exceedances)
```



การหาค่าประมาณพารามิเตอร์ที่เหมาะสม

```
A = gpd.fit(weather, u)
$threshold
[1] 80
$nexc
[1] 199

$conv
[1] 0
$nllh
[1] 710.2624

$mle
[1] 15.6139392 -0.1787775

$rate
[1] 0.009877891
$se
[1] 1.41400151 0.05777677

gpd.diag(A)
```



AIC & BIC

```
aic.A = -2 * A$nllh + 2 * 2
bic.A = -2 * A$nllh + log(length(A$data)) * length(A$mle)
fit1  <- c(par = A$mle,
           se = A$se,
           aic = aic.A,
           bic = bic.A)
fit1
        par1          par2          se1          se2
 1.561394e+01 -1.787775e-01  1.414002e+00  5.777677e-02
         aic           bic
-1.416525e+03 -1.409938e+03
aic.A
[1] -1416.525
bic.A
[1] -1409.938
```

GOF test

```
loc = 0
ks.test(threshold.exceedances, "pgpd", loc, A$mle[1], A$mle[2])

        One-sample Kolmogorov-Smirnov test

data:  threshold.exceedances

D = 0.07206, p-value = 0.2527

alternative hypothesis: two-sided

Warning message:

In ks.test(threshold.exceedances, "pgpd", loc, A$mle[1], A$mle[2]) :
  ties should not be present for the Kolmogorov-Smirnov test
```

return level using

```
AA <- fevd(weather,
        threshold = u,
        type      = "GP",
        method    = "MLE")
return.level(AA, return.period = c(2, 5, 10, 50, 100))

fevd(x = weather, threshold = u, type = "GP", method = "MLE")
get(paste("return.level.fevd.", newcl, sep = ""))(x = x,
    return.period = return.period)

 GP model fitted to  weather
Data are assumed to be  stationary
[1] "Return Levels for period units in years"
  2-year level   5-year level  10-year level  50-year level  100-year level
      105.9907       115.2559       121.3220       132.8173     136.8352
```

using rlevd function

```
rlevd(c(2, 10, 50, 100),
      loc   = 0,
      scale = A$mle[1],
      shape = A$mle[2],
      u,
      type  = "GP",
      rate  = A$rate)

       2       10       50      100
105.9953 121.3331 132.8358 136.8569
```

Dindaeng Station PM2.5

การเรียกใช้แพคเกจของโปรแกรม

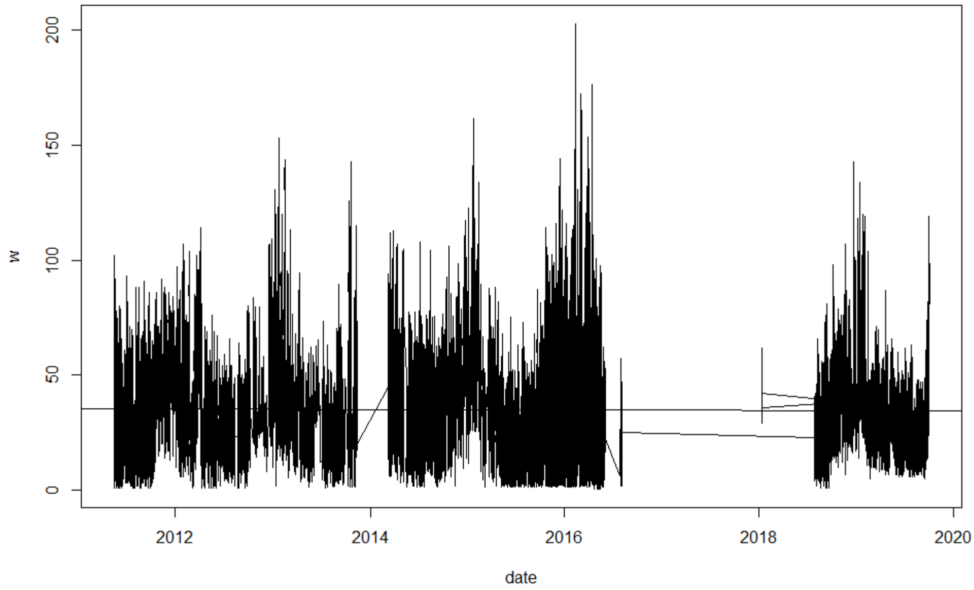
```
library(extRemes)
library(ismev)
library(aTSA)
library(tseries)
library(trend)
library(e1071)
library(TLMoments)
library(nortest)
library(eva)
library(ADGofTest)
library(lattice)
```

การเรียกใช้ข้อมูล

```
data <-
read.csv("C:\\Users\\acer\\Desktop\\Project\\Code\\clean_Dindaeng\\clean_DinDaeng
_pm2.5_hour.csv", header = T)
```

การสร้างกราฟเพื่อดูแนวโน้ม

```
date <- as.POSIXct(data$DAY, format = "%d/%m/%Y %H:%M")
linearmodel <- lm(data$PM2.5 ~ date)
plot(data$PM2.5 ~ date, type = 'l')
abline(linearmodel)
weather = na.omit(data$PM2.5)
```



การทดสอบเพื่อดูแนวโน้มของข้อมูล

mk.test(weather, continuity = T)

Mann-Kendall trend test

data: weather

z = -3.6425, n = 47913, p-value = 0.0002701

alternative hypothesis: true S is not equal to 0

sample estimates:

```
            S          varS           tau
-1.273358e+07  1.222118e+13 -1.112154e-02
```

หา summary

summary(weather)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.74   20.93   31.28   35.06   45.00  202.64
```

การทดสอบความเป็น Stationary ความเบ้ ความโค้ง

stationary.test(weather)

Augmented Dickey-Fuller Test

alternative: stationary

Type 1: no drift no trend

```
      lag    ADF p.value
 [1,]   0 -29.99    0.01
 [2,]   1 -37.06    0.01
 [3,]   2 -31.23    0.01
 [4,]   3 -23.86    0.01
 [5,]   4 -18.56    0.01
 [6,]   5 -16.59    0.01
 [7,]   6 -15.00    0.01
 [8,]   7 -15.17    0.01
 [9,]   8 -15.09    0.01
[10,]   9 -13.14    0.01
[11,]  10 -11.58    0.01
[12,]  11 -10.19    0.01
[13,]  12 -10.07    0.01
[14,]  13  -9.67    0.01
[15,]  14  -9.50    0.01
[16,]  15  -9.88    0.01
```

Type 2: with drift no trend

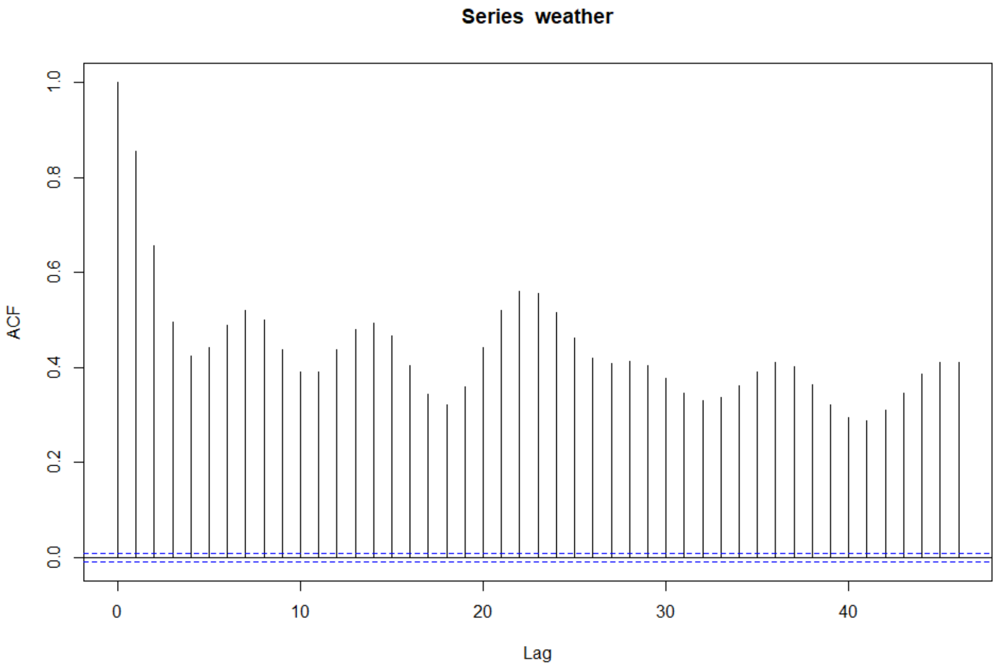
```
      lag   ADF p.value
 [1,]   0 -61.0    0.01
 [2,]   1 -78.9    0.01
 [3,]   2 -69.0    0.01
 [4,]   3 -54.1    0.01
 [5,]   4 -42.7    0.01
 [6,]   5 -38.7    0.01
 [7,]   6 -35.4    0.01
 [8,]   7 -36.2    0.01
 [9,]   8 -36.4    0.01
```

```
[10,]   9 -32.0    0.01
[11,]  10 -28.4    0.01
[12,]  11 -25.1    0.01
[13,]  12 -25.0    0.01
[14,]  13 -24.1    0.01
[15,]  14 -23.8    0.01
[16,]  15 -24.8    0.01
Type 3: with drift and trend
      lag   ADF p.value
 [1,]   0 -61.0    0.01
 [2,]   1 -78.9    0.01
 [3,]   2 -69.0    0.01
 [4,]   3 -54.1    0.01
 [5,]   4 -42.7    0.01
 [6,]   5 -38.7    0.01
 [7,]   6 -35.4    0.01
 [8,]   7 -36.2    0.01
 [9,]   8 -36.4    0.01
[10,]   9 -32.0    0.01
[11,]  10 -28.4    0.01
[12,]  11 -25.1    0.01
[13,]  12 -25.0    0.01
[14,]  13 -24.1    0.01
[15,]  14 -23.8    0.01
[16,]  15 -24.8    0.01
----
Note: in fact, p.value = 0.01 means p.value <= 0.01
skewness(weather)
[1] 1.239852
```

kurtosis(weather)

[1] 2.427285

acf(weather)



adf.test(weather)

Augmented Dickey-Fuller Test

data: weather

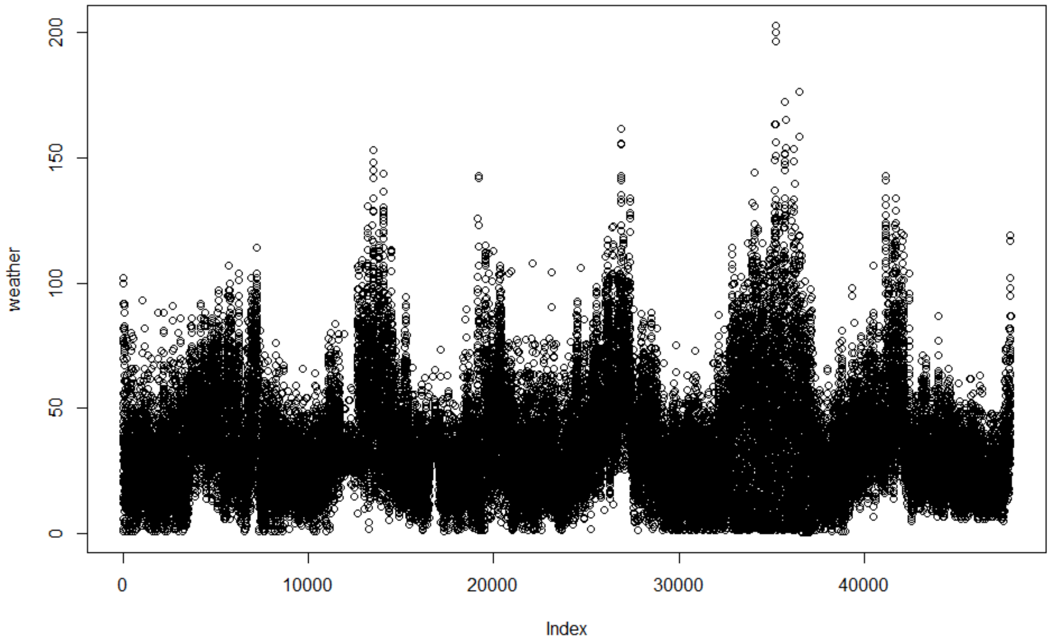
Dickey-Fuller = -14.514, Lag order = 36, p-value =

0.01

alternative hypothesis: stationary

การสร้างกราฟเพื่อดูการกระจายตัวของข้อมูล

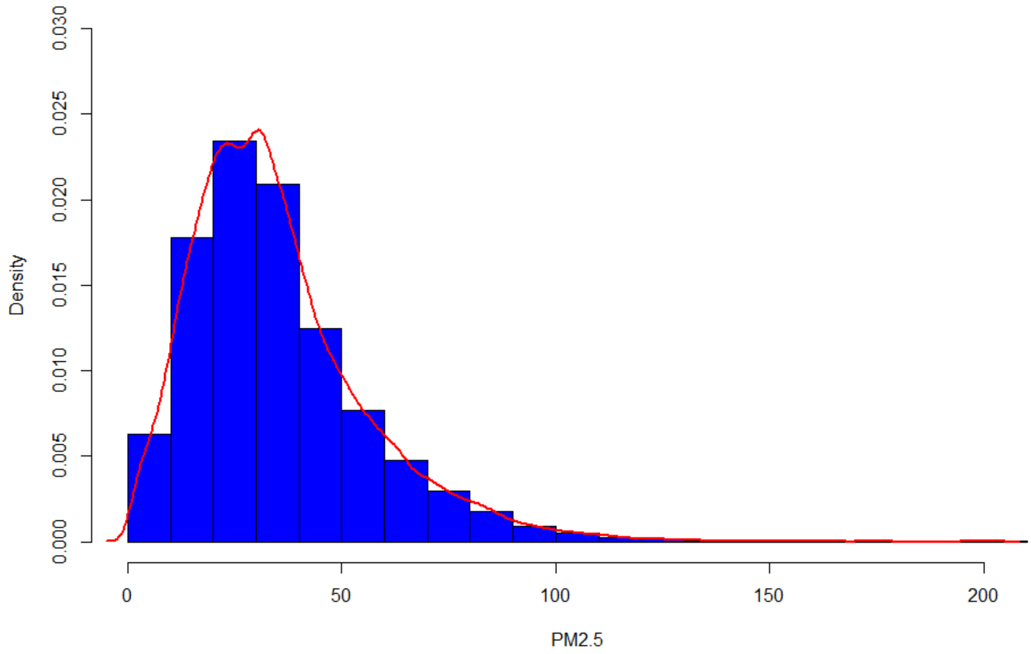
```
plot(weather)
```



การสร้างแผนภูมิความถี่

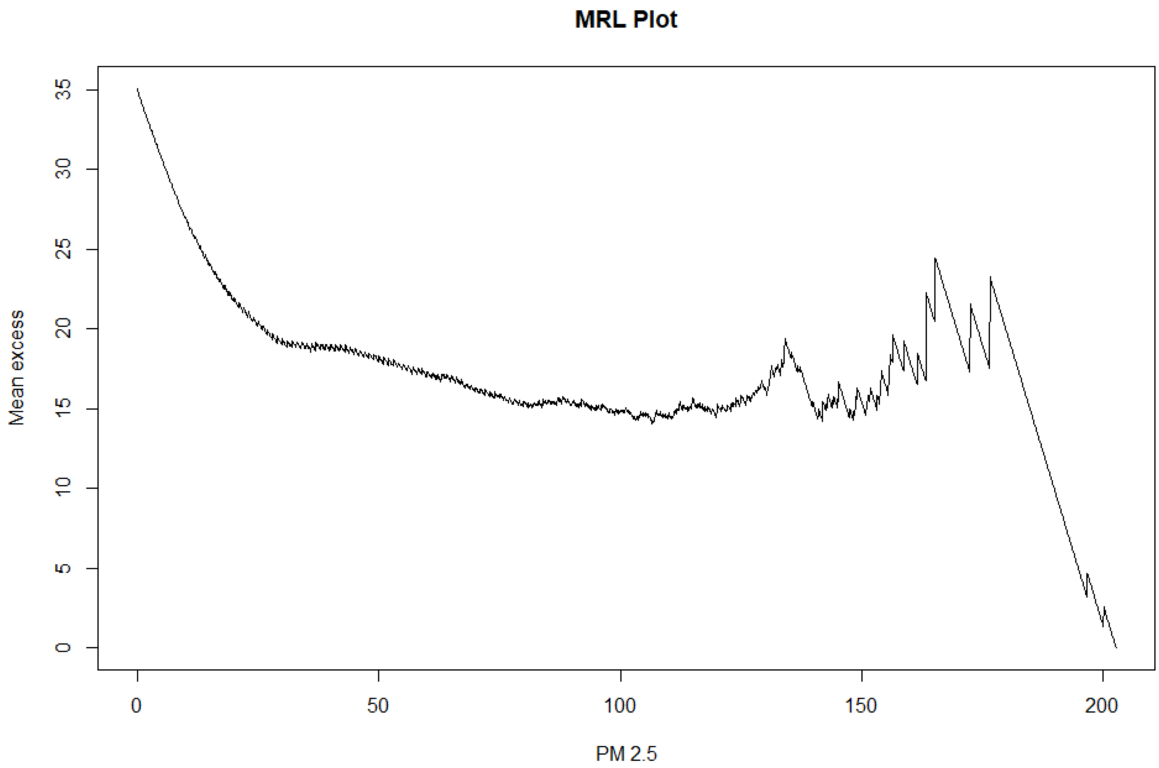
```
par(mfrow = c(1,1))
plot.ts(weather,
        xlab = "Days",
        ylab = "original data")
hist(weather,
     xlab  = "PM2.5",
     ylim  = c(0, 0.04),
     main  = " ",
     col   = "blue",
     border = "black",
     prob  = TRUE)
lines(density(weather),
      xlab  = " ",
      main = " ",
```

```
      col   = "red",
      lty   = 1,
      lwd  = 2)
```



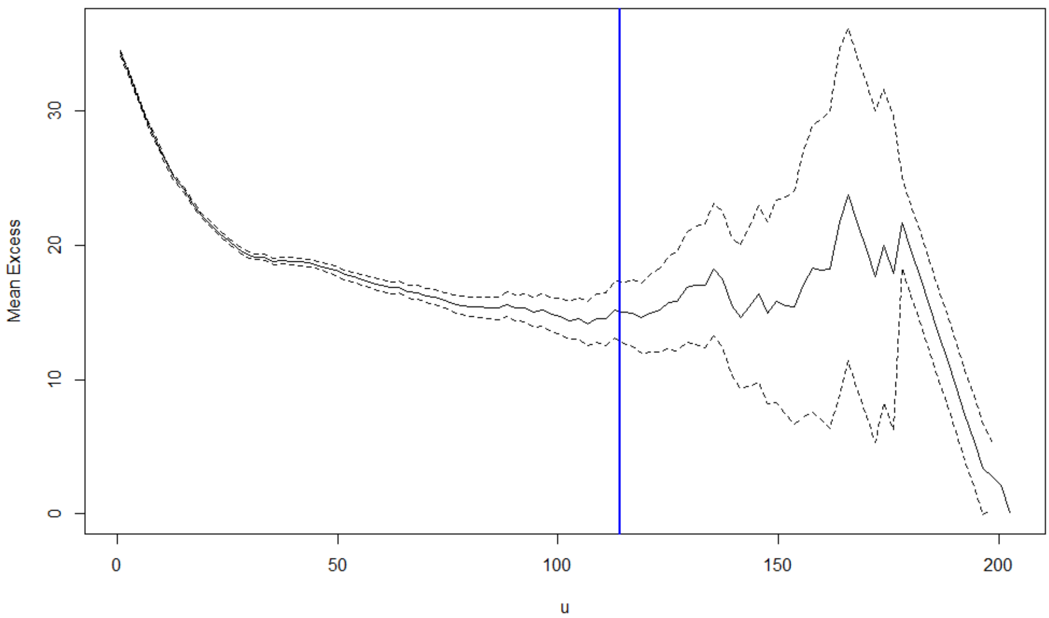
การสร้างกราฟ MRL

```
u <- seq (0, max(weather), 0.1)
x <- vector("numeric", length(u))

for (i in 1 : length (x) ) {
  threshold.exceedances <- weather[weather > u[i]]
  x[i] <- mean(threshold.exceedances - u[i])
}
plot(x ~ u,
     xlab = "PM 2.5",
     ylab = "Mean excess",
     main = "MRL Plot",
     type = "l")
```



การกำหนดค่าเกณฑ์บนกราฟ MRL

```
mrl.plot(weather)
abline(v = 114,
       col = "blue",
       lwd = 2)
```



หาค่า Percentile

```
duration = data$PM2.5
quantile(duration, c( .95, .96, .97, .98, .99, .995, .996))
   95%    96%    97%    98%    99%    99.5%   99.6%
 75.03  79.00  83.57  89.78  100.65  111.00  114.00
```

Data pre-processing

```
u <- 114
above.threshold <- weather[weather > u]
length(above.threshold)
[1] 189
threshold.exceedances <- above.threshold - u
above.threshold
  [1] 118.00 121.88 130.68 119.81 117.48 123.17 128.35 133.79
  [9] 141.81 145.17 153.20 148.28 129.13 116.44 114.89 119.81
 [17] 114.37 128.35 143.88 136.63 124.98 124.98 120.84 114.37
 [25] 117.74 126.28 127.57 120.07 128.87 130.42 123.43 123.43
 [33] 114.37 126.02 114.63 142.84 141.81 123.17 114.89 114.89
 [41] 117.48 122.14 122.65 115.67 114.63 121.10 122.91 133.27
 [49] 141.03 131.97 135.34 142.59 161.48 155.79 155.53 141.81
 [57] 120.84 114.89 118.25 117.48 115.92 127.83 126.02 127.05
 [65] 132.49 134.04 120.58 114.37 116.18 114.63 116.44 131.97
 [73] 120.07 123.69 117.48 144.14 130.94 121.10 119.03 120.84
 [81] 121.62 115.92 114.89 120.07 118.00 127.83 131.20 116.44
 [89] 124.73 125.76 137.15 163.29 149.06 121.36 120.07 120.07
 [97] 133.27 163.29 150.87 123.95 118.25 196.69 200.05 202.64
[105] 156.31 126.28 121.10 124.98 130.68 115.67 115.92 128.09
[113] 119.81 124.73 130.94 125.24 122.40 114.89 127.05 172.35
[121] 151.65 122.14 145.17 147.50 131.20 134.04 127.31 129.90
```

[129] 124.21 126.02 148.54 144.92 126.54 151.13 165.11 153.98

[137] 127.31 118.77 120.32 129.13 132.23 124.73 115.67 133.27

[145] 130.42 153.72 148.80 114.11 128.61 139.74 116.18 122.14

[153] 116.44 176.50 158.63 119.29 118.51 114.63 124.00 127.00

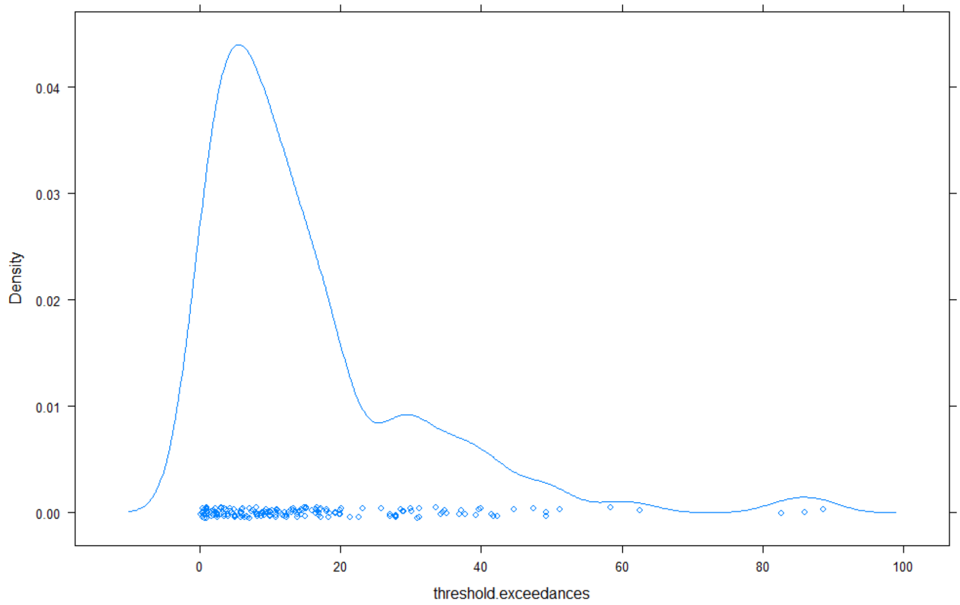
[161] 117.00 122.00 131.00 123.00 124.00 120.00 118.00 123.00

[169] 143.00 141.00 134.00 119.00 118.00 115.00 115.00 123.00

[177] 116.00 119.00 122.00 124.00 115.00 117.00 129.00 134.00

[185] 120.00 117.00 119.00 119.00 117.00

densityplot(threshold.exceedances)



การหาค่าประมาณพารามิเตอร์ที่เหมาะสม

A = gpd.fit(weather, u)

$threshold

[1] 114

$nexc

[1] 189

$conv

[1] 0

$nllh

[1] 702.5608

$mle

[1] 14.63642754 0.03388451

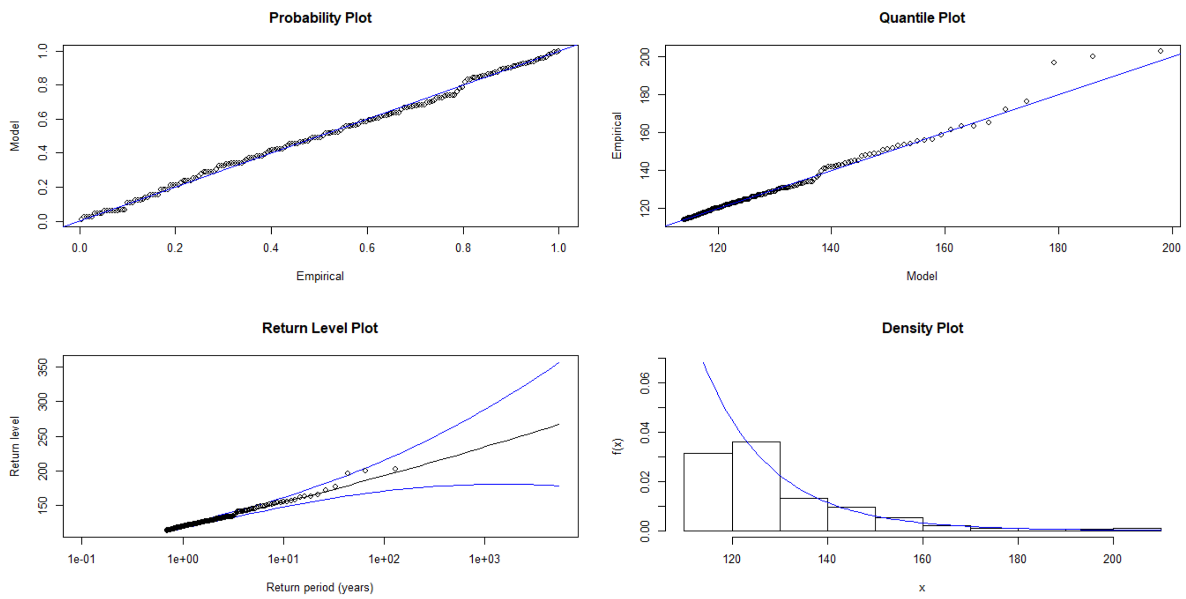
$rate

[1] 0.00394465

$se

[1] 1.54414421 0.07646826

gpd.diag(A)



AIC & BIC

aic.A = -2 * A$nllh + 2 * 2

bic.A = -2 * A$nllh + log(length(A$data)) * length(A$mle)

fit1 <- c(par = A$mle,

```
            se = A$se,
            aic = aic.A,
            bic = bic.A)
fit1
         par1          par2           se1           se2
 1.463643e+01  3.388451e-02  1.544144e+00  7.646826e-02
          aic           bic
-1.401122e+03  -1.394638e+03
aic.A
[1] -1401.122
bic.A
[1] -1394.638
                                             # GOF test #
loc = 0
ks.test(threshold.exceedances, "pgpd", loc, A$mle[1], A$mle[2])


          One-sample Kolmogorov-Smirnov test


data:  threshold.exceedances
D = 0.04534, p-value = 0.832
alternative hypothesis: two-sided


Warning message:
In ks.test(threshold.exceedances, "pgpd", loc, A$mle[1], A$mle[2]) :
  ties should not be present for the Kolmogorov-Smirnov test
```

return level using

```
AA <- fevd(weather,
        threshold = u,
        type      = "GP",
        method    = "MLE")
return.level(AA, return.period = c(2, 5, 10, 50, 100))

fevd(x = weather, threshold = u, type = "GP", method = "MLE")
get(paste("return.level.fevd.", newcl, sep = ""))(x = x,
   return.period = return.period)

 GP model fitted to  weather
Data are assumed to be  stationary
[1] "Return Levels for period units in years"
  2-year level   5-year level  10-year level  50-year level
      129.7639       143.8806       154.8572       181.3702
100-year level
      193.2460
```

using rlevd function

```
rlevd(c(2, 10, 50, 100),
      loc   = 0,
      scale = A$mle[1],
      shape = A$mle[2],
      u,
      type  = "GP",
      rate  = A$rate)

       2       10       50      100
129.7713 154.8659 181.3670 193.2332
```

Jula Pathumwan PM2.5

การเรียกใช้แพคเกจของโปรแกรม

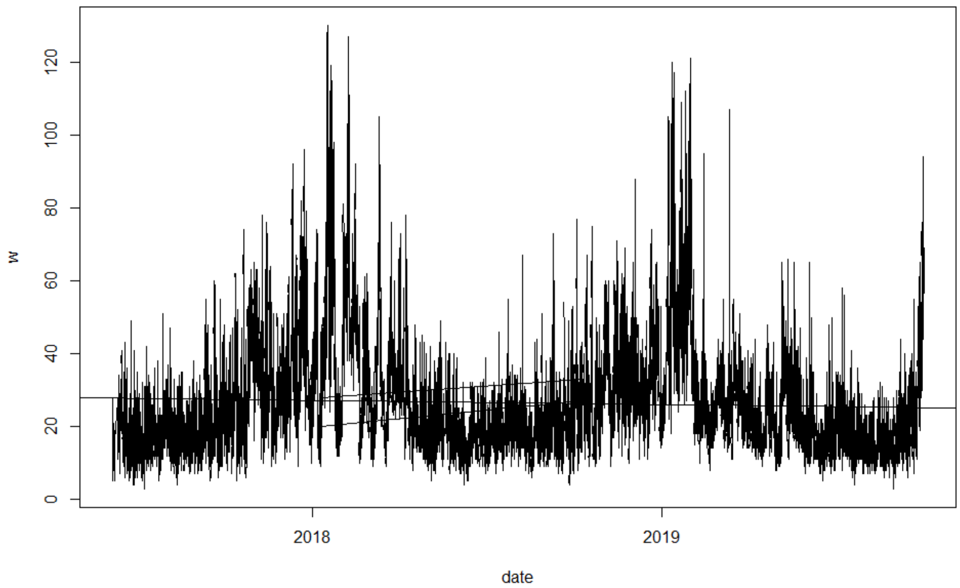
```
library(extRemes)
library(ismev)
library(aTSA)
library(tseries)
library(trend)
library(e1071)
library(TLMoments)
library(nortest)
library(eva)
library(ADGofTest)
library(lattice)
```

การเรียกใช้ข้อมูล

```
data <-
read.csv("C:\\Users\\acer\\Desktop\\Project\\Code\\clean_Jula_Pathumwan\\clean_
Jula_Pathumwan_pm2.5_hour.csv", header = T)
```

การสร้างกราฟเพื่อดูแนวโน้ม

```
date <- as.POSIXct(data$DAY, format = "%d/%m/%Y %H:%M")
linearmodel <- lm(data$PM2.5 ~ date)
plot(data$PM2.5 ~ date, type = 'l')
abline(linearmodel)
weather = na.omit(data$PM2.5)
```



การทดสอบเพื่อดูแนวโน้มของข้อมูล

mk.test(weather, continuity = T)

Mann-Kendall trend test

data: weather

z = -6.4776, n = 20105, p-value = 9.322e-11

alternative hypothesis: true S is not equal to 0

sample estimates:

S	varS	tau
-6.152671e+06	9.022020e+11	-3.086315e-02

หา summary

summary(weather)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3.00	16.00	22.00	26.54	32.00	130.00

การทดสอบความเป็น Stationary ความเบ้ ความโค้ง

stationary.test(weather)

Augmented Dickey-Fuller Test

alternative: stationary

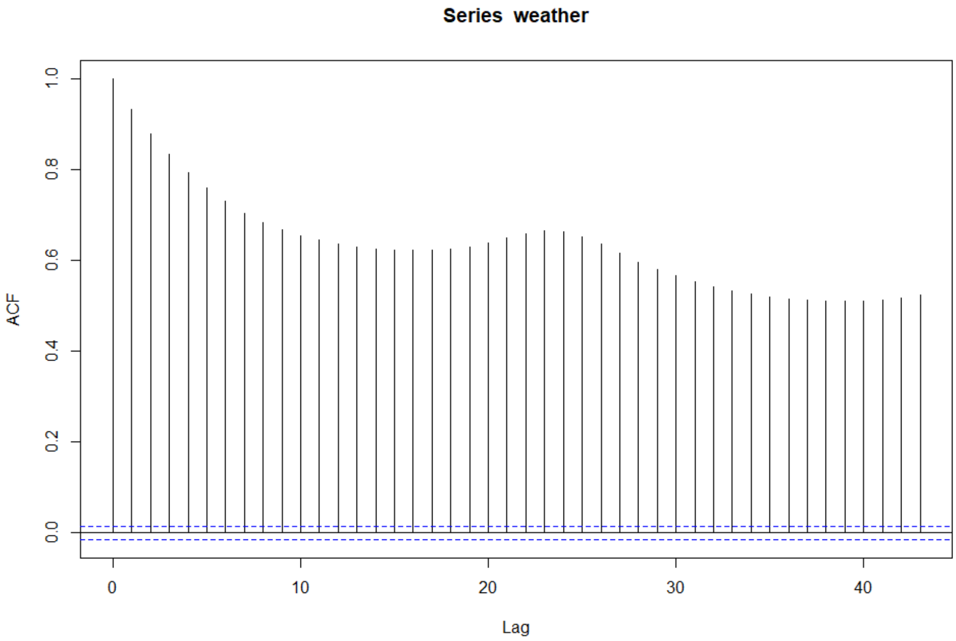
Type 1: no drift no trend

	lag	ADF	p.value
[1,]	0	-12.79	0.01
[2,]	1	-11.58	0.01
[3,]	2	-10.81	0.01
[4,]	3	-10.28	0.01
[5,]	4	-9.68	0.01
[6,]	5	-9.17	0.01
[7,]	6	-8.75	0.01
[8,]	7	-8.28	0.01
[9,]	8	-7.72	0.01
[10,]	9	-7.28	0.01
[11,]	10	-6.96	0.01
[12,]	11	-6.63	0.01
[13,]	12	-6.30	0.01

Type 2: with drift no trend

	lag	ADF	p.value
[1,]	0	-26.3	0.01
[2,]	1	-24.1	0.01
[3,]	2	-22.8	0.01
[4,]	3	-21.9	0.01
[5,]	4	-20.8	0.01
[6,]	5	-19.8	0.01
[7,]	6	-19.1	0.01
[8,]	7	-18.2	0.01
[9,]	8	-17.1	0.01
[10,]	9	-16.2	0.01
[11,]	10	-15.6	0.01
[12,]	11	-15.0	0.01

```
[13,]  12 -14.3    0.01
Type 3: with drift and trend
      lag    ADF p.value
 [1,]   0 -26.4    0.01
 [2,]   1 -24.2    0.01
 [3,]   2 -22.8    0.01
 [4,]   3 -21.9    0.01
 [5,]   4 -20.8    0.01
 [6,]   5 -19.9    0.01
 [7,]   6 -19.1    0.01
 [8,]   7 -18.2    0.01
 [9,]   8 -17.1    0.01
[10,]   9 -16.2    0.01
[11,]  10 -15.6    0.01
[12,]  11 -15.0    0.01
[13,]  12 -14.3    0.01
----
Note: in fact, p.value = 0.01 means p.value <= 0.01
skewness(weather)
[1] 1.839202
kurtosis(weather)
[1] 4.858087
acf(weather)
```



adf.test(weather)

Augmented Dickey-Fuller Test

data: weather

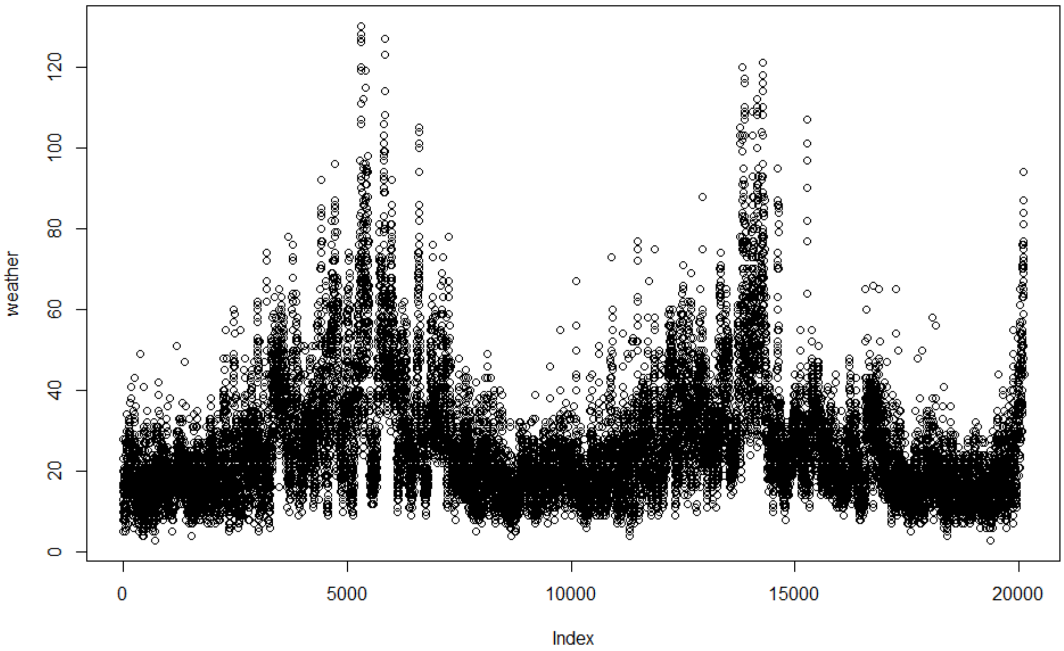
Dickey-Fuller = -9.7604, Lag order = 27, p-value =

0.01

alternative hypothesis: stationary

การสร้างกราฟเพื่อดูการกระจายตัวของข้อมูล

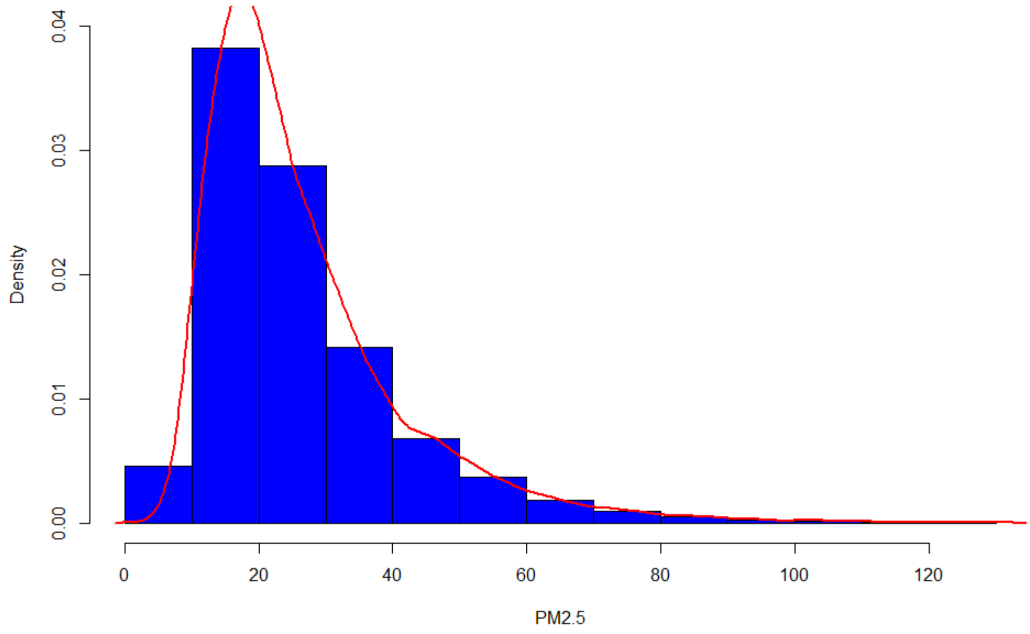
```
plot(weather)
```



การสร้างแผนภูมิความถี่

```
par(mfrow = c(1,1))
plot.ts(weather,
        xlab = "Days",
        ylab = "original data")
hist(weather,
     xlab  = "PM2.5",
     ylim  = c(0, 0.04),
     main  = " ",
     col   = "blue",
     border = "black",
     prob  = TRUE)
lines(density(weather),
```

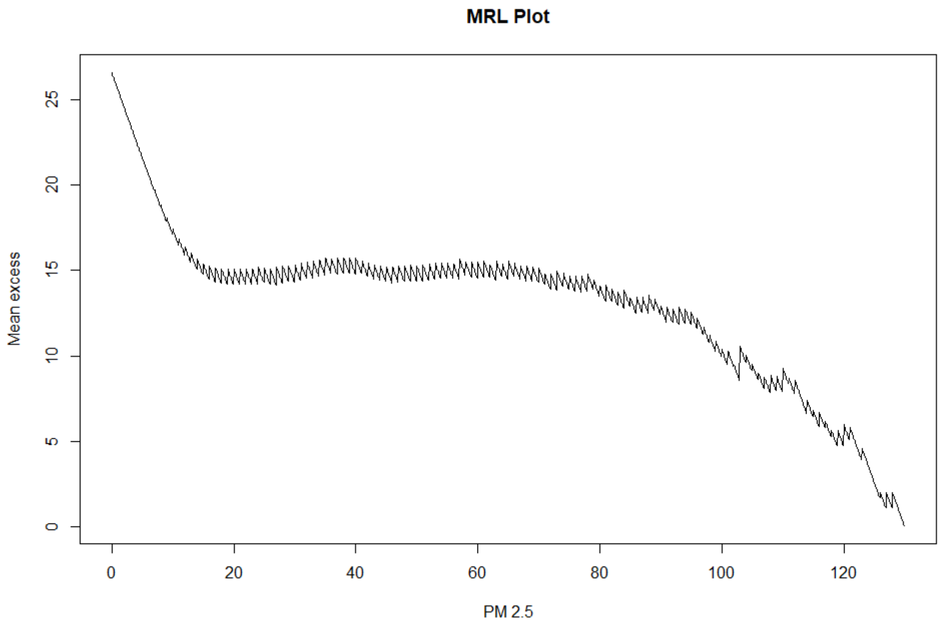
```
      xlab  = " ",
      main = " ",
      col   = "red",
      lty    = 1,
      lwd  = 2)
```



การสร้างกราฟ MRL

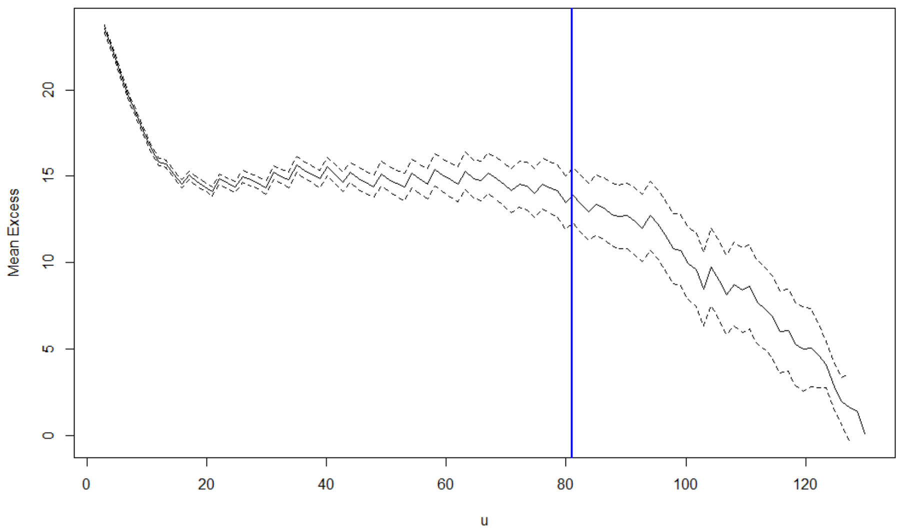
```
u <- seq (0, max(weather), 0.1)
x <- vector("numeric", length(u))

for (i in 1 : length (x) ) {
  threshold.exceedances <- weather[weather > u[i]]
  x[i] <- mean(threshold.exceedances - u[i])
}
plot(x ~ u,
     xlab = "PM 2.5",
     ylab = "Mean excess",
     main = "MRL Plot",
```

```
type = "l")
```



การกำหนดค่าเกณฑ์บนกราฟ MRL

```
mrl.plot(weather)
abline(v = 81,
       col = "blue",
       lwd = 2)
```

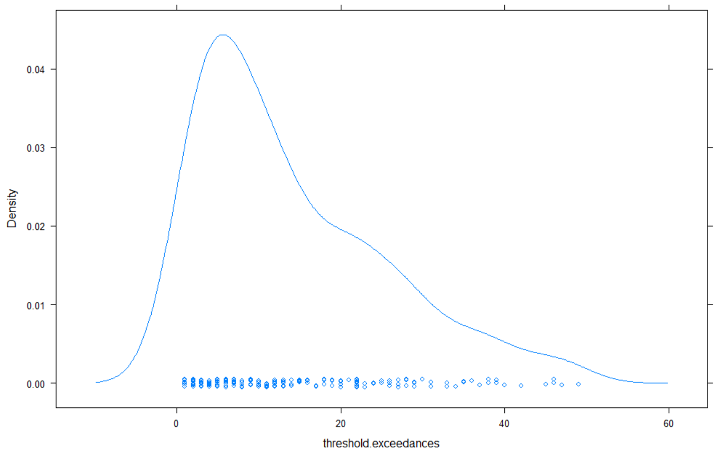


หาค่า Percentile

```
duration = data$PM2.5
quantile(duration, c( .95, .96, .97, .98, .99))
95%  96%  97%  98%  99%
 56   60    64    70    81
```

Data pre-processing

```
u <- 81
above.threshold <- weather[weather > u]
length(above.threshold)
[1] 196
threshold.exceedances <- above.threshold - u
above.threshold
 [1]  84  85  85  84  83  92  84  85  82  86  86  82  86  85
 [15]  87  96  84  83  97 106 107 128 119 130 120 126  92 111
 [29] 127  90  93  84  89  82 112  86  96  87  95  86 119  83
 [43]  82  83  91  88  87  94  96 115  93  95  94  91  85  91
 [57]  88  94  98  84  82  83  92  89  99  98  99  97  98  93
 [71] 101 106  89  90  90 103  94  99 127 123 114 108  89  84
 [85]  86  92  86  94 101 100 101 104 105 100  84  82  82  88
 [99] 105 103 101  83  99 103 102  84  91 103 120  91 103  87
[113]  82  88  89  92  95  92  83  82  91 107 110 103 103 117
[127] 109 116 108  86  86  88  93 103 109  93  83  83  88  88
[141]  82  89  91  87  84 100 108 112 109 110  90  87  88  83
[155]  86  91  87  95  93  88  93 104  88  92  84  89 114 108
[169] 110 103 116 110 118 121 103  93  96  83  85  88  84  87
[183]  86  95  87  84  86  85  82 101 107  97  90  87  94  84
densityplot(threshold.exceedances)
```



การหาค่าประมาณพารามิเตอร์ที่เหมาะสม

A = gpd.fit(weather, u)

$threshold

[1] 81

$nexc

[1] 196

$conv

[1] 0

$nllh

[1] 708.6773

$mle

[1] 18.4287776 -0.2983433

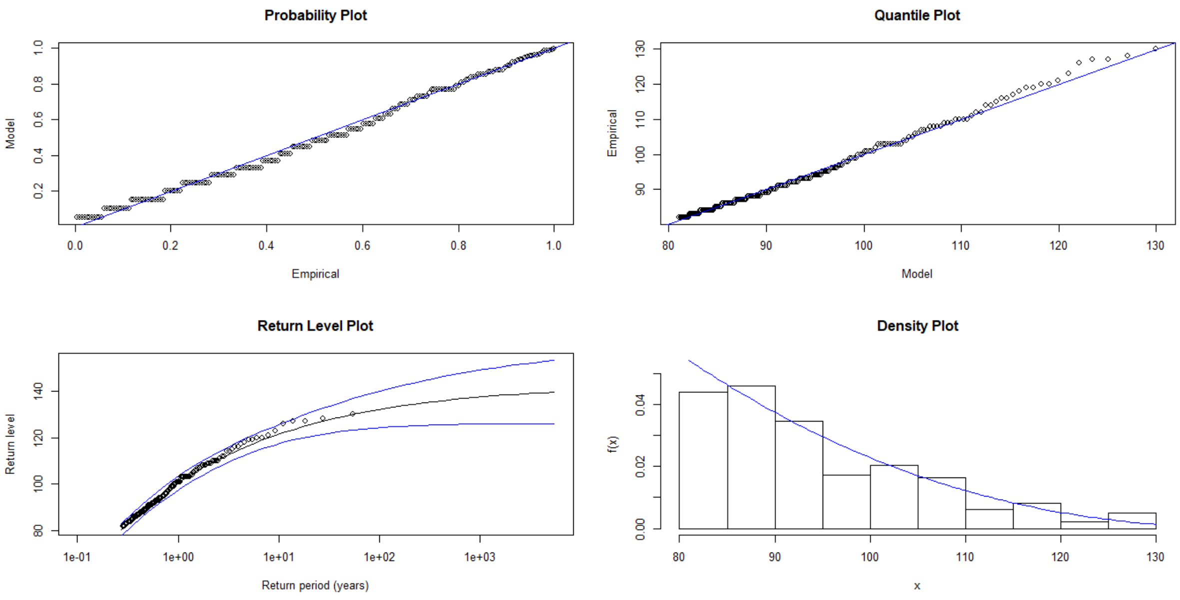
$rate

[1] 0.009748819

$se

[1] 1.78616165 0.06889261

gpd.diag(A)




```
# AIC & BIC #

aic.A = -2 * A$nllh + 2 * 2

bic.A = -2 * A$nllh + log(length(A$data)) * length(A$mle)

fit1  <- c(par = A$mle,
           se = A$se,
           aic = aic.A,
           bic = bic.A)

fit1

        par1          par2          se1          se2
 1.842878e+01 -2.983433e-01  1.786162e+00  6.889261e-02
         aic           bic
-1.413355e+03 -1.406798e+03

aic.A

[1] -1413.355

bic.A

[1] -1406.798
```

```
# GOF test #

loc = 0

ks.test(threshold.exceedances, "pgpd", loc, A$mle[1], A$mle[2])
```

```
        One-sample Kolmogorov-Smirnov test

data:  threshold.exceedances
D = 0.055973, p-value = 0.571
alternative hypothesis: two-sided

Warning message:
In ks.test(threshold.exceedances, "pgpd", loc, A$mle[1], A$mle[2]) :
  ties should not be present for the Kolmogorov-Smirnov test
```

```
                              # return level using #
AA <- fevd(weather,
          threshold = u,
          type      = "GP",
          method    = "MLE")
return.level(AA, return.period = c(2, 5, 10, 50, 100))

fevd(x = weather, threshold = u, type = "GP", method = "MLE")
get(paste("return.level.fevd.", newcl, sep = ""))(x = x,
    return.period = return.period)

 GP model fitted to  weather
Data are assumed to be  stationary
[1] "Return Levels for period units in years"
  2-year level   5-year level  10-year level  50-year level
      108.3851       116.6103       121.4974       129.6083
100-year level
      132.0663
```

using rlevd function

```
rlevd(c(2, 10, 50, 100),
      loc   = 0,
      scale = A$mle[1],
      shape = A$mle[2],
      u,
      type  = "GP",
      rate  = A$rate)

      2        10        50       100
108.3813 121.4945 129.6074 132.0664
```

ตัวแบบ GEV รายสัปดาห์

CHOKCHAI Metropolitan Police Station WangThonglang PM2.5

การเรียกใช้แพคเกจของโปรแกรม

```
library(extRemes)
library(ismev)
library(aTSA)
library(tseries)
library(trend)
library(e1071)
library(TLMoments)
library(nortest)
library(eva)
library(ADGofTest)
library(lattice)
```

การเรียกใช้ข้อมูล

```
data <-
read.csv("C:\\Users\\acer\\Desktop\\Project\\Code\\clean_ChokchaiMetropolitan
PoliceStation_WangThonglang\\clean_ChokchaiMetropolitanPoliceStation_WangThong
lang_pm2.5_week.csv", header = T)
```

การจัดการข้อมูล และ หา summary

```
weather = na.omit(data$PM2.5)
summary(weather)
  Min.  1st Qu.  Median   Mean  3rd Qu.    Max.
  19.00  37.25    46.50   53.64   61.75   138.00
```

การทดสอบเพื่อดูแนวโน้มของข้อมูล

mk.test(weather, continuity = T)

Mann-Kendall trend test

data: weather

z = 0.21474, n = 122, p-value = 0.83

alternative hypothesis: true S is not equal to 0

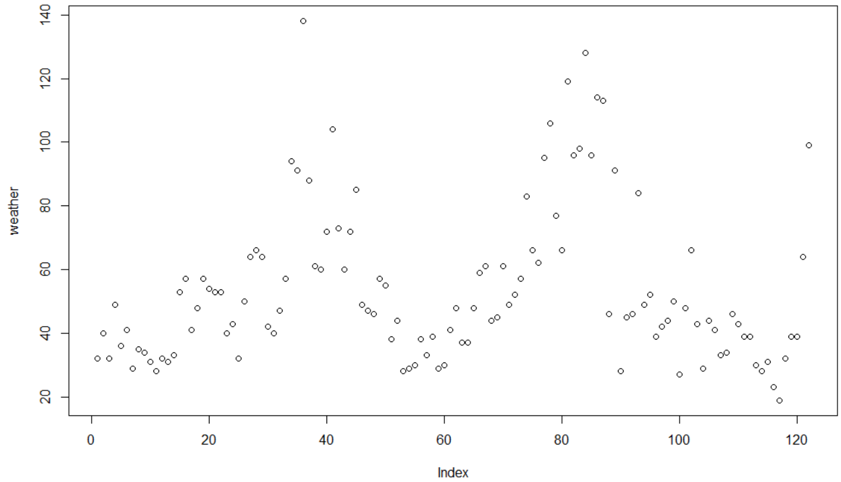
sample estimates:

S varS tau

9.800000e+01 2.040353e+05 1.338568e-02

การสร้างกราฟเพื่อดูความหนาแน่นของข้อมูล

plot(weather)



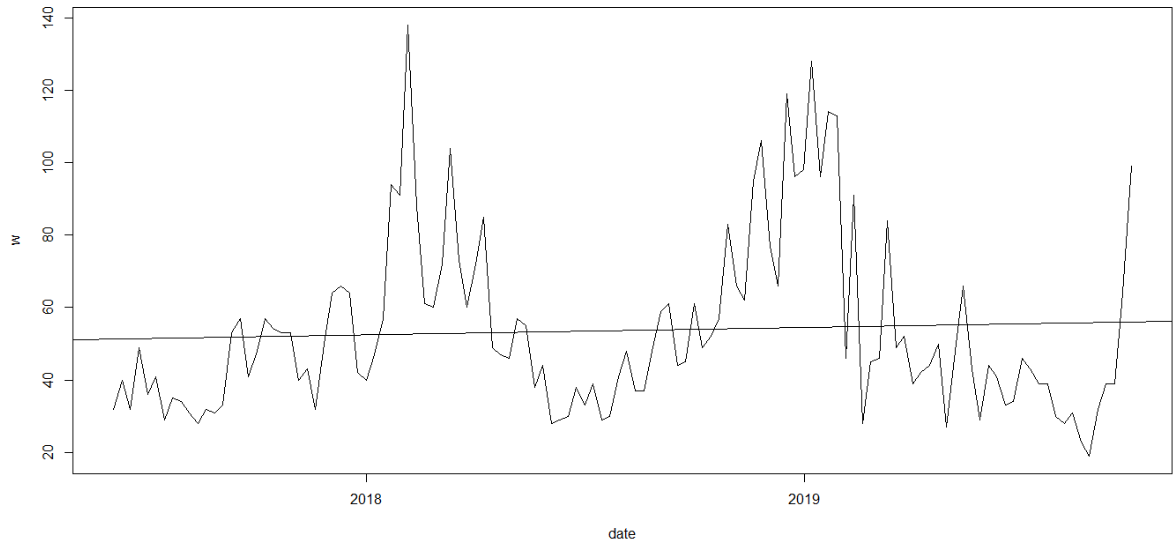
การสร้างกราฟเพื่อดูแนวโน้ม

date <- as.Date(as.character(data$DAY), format = "%Y-%m-%d")

linearmodel <- lm(data$PM2.5 ~ date)

plot(data$PM2.5 ~ date, type = 'l')

abline(linearmodel)



การทดสอบความเป็น Stationary ความเบ้ ความโค้ง

stationary.test(weather)

Augmented Dickey-Fuller Test

alternative: stationary

Type 1: no drift no trend

```
     lag    ADF p.value
[1,]   0 -1.423   0.168
[2,]   1 -0.886   0.361
[3,]   2 -0.654   0.445
[4,]   3 -0.587   0.469
[5,]   4 -0.532   0.489
```

Type 2: with drift no trend

```
     lag   ADF p.value
[1,]   0 -4.48  0.0100
[2,]   1 -3.28  0.0196
[3,]   2 -2.85  0.0569
[4,]   3 -2.55  0.1155
[5,]   4 -2.55  0.1167
```

Type 3: with drift and trend

```
     lag   ADF p.value
[1,]   0 -4.47  0.0100
[2,]   1 -3.28  0.0783
[3,]   2 -2.85  0.2233
[4,]   3 -2.54  0.3477
[5,]   4 -2.54  0.3497
----
Note: in fact, p.value = 0.01 means p.value <= 0.01
```

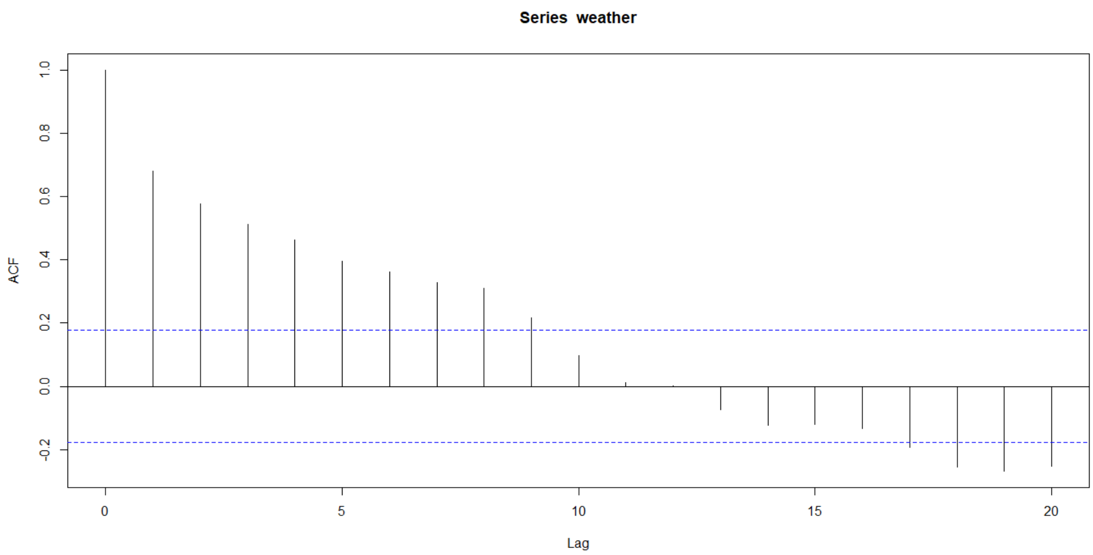
skewness(weather)

[1] 1.300627

kurtosis(weather)

[1] 1.209228

acf(weather)



adf.test(weather)

Augmented Dickey-Fuller Test

data: weather

```
Dickey-Fuller = -2.5381, Lag order = 4, p-value =
0.353
alternative hypothesis: stationary
```

การสร้างโมเดล หาค่าประมาณพารามิเตอร์ที่เหมาะสม AIC&BIC และ ระดับการเกิดซ้ำ

```
dat <- read.csv("C:\\Users\\acer\\Desktop\\Project\\Code\\clean_Chokchai
MetropolitanPoliceStation_WangThonglang\\clean_ChokchaiMetropolitanPoliceStatio
n_WangThonglang_pm2.5_week.csv", header = T)
length(dat)
n = nrow(dat)
dat_ = dat[,2]

## Model1 ##
model1 <- gev.fit(dat_)
$conv
[1] 0
$nllh
[1] 535.8281
$mle
[1] 41.4400835 14.6625648  0.2199876
$se
[1] 1.51426319 1.22357176 0.07878155

aic.model1 = -2*model1$nllh+2*3
aic.model1
[1] -1065.656

bic.model1 = -2*model1$nllh+log(length(dat_))*length(model1$mle)
bic.model1
```

```
[1] -1057.244

fit1  <- c(par = model1$mle, se = model1$se, aic = aic.model1, bic = bic.model1)
fit1
        par1           par2          par3           se1
 4.144008e+01  1.466256e+01  2.199876e-01  1.514263e+00
          se2           se3           aic           bic
 1.223572e+00  7.878155e-02 -1.065656e+03 -1.057244e+03
```

การหาความเหมาะสมของตัวแบบ

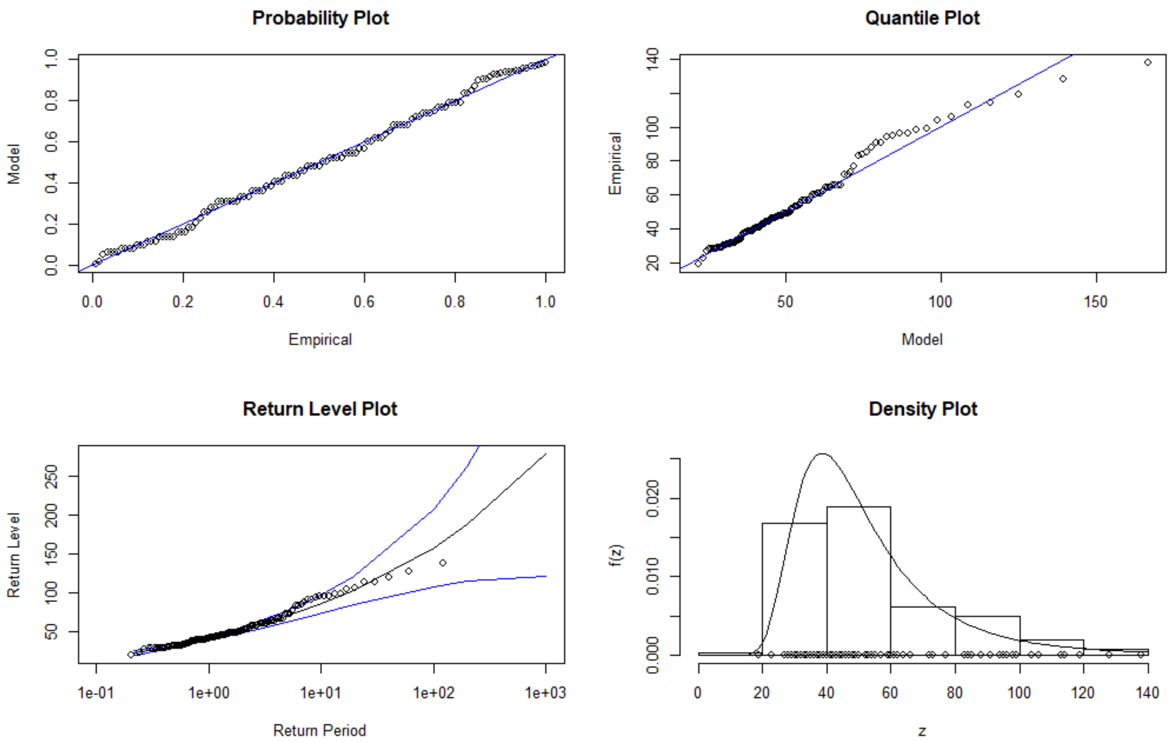
```
### GOF TEST ###
ks.test(dat$PM2.5, pgev, model1$mle[1], model1$mle[2], model1$mle[3])

One-sample Kolmogorov-Smirnov test

data:  dat$PM2.5
D = 0.051139, p-value = 0.9071
alternative hypothesis: two-sided

Warning message:
In ks.test(dat$PM2.5, pgev, model1$mle[1], model1$mle[2], model1$mle[3]) :
 ties should not be present for the Kolmogorov-Smirnov test

A = fevd(dat_, threshold = NULL, threshold.fun = ~1, location.fun = ~1,
         scale.fun = ~1, shape.fun = ~1, use.phi = FALSE,
         type = "GEV", method = "MLE", initial = NULL,
         units = NULL, time.units = "weeks", period.basis = "year",
         optim.args = NULL, priorFun = NULL,
         priorParams = NULL, proposalFun = NULL, proposalParams = NULL,
         iter = 9999, weights = 1, blocks = NULL, verbose = FALSE)
```

```
A
[1] "Estimation Method used: MLE"

 Negative Log-Likelihood Value:  535.8281

 Estimated parameters:
  location      scale      shape
41.4437529 14.6623602  0.2202806



 Standard Error Estimates:
  location      scale      shape
1.51441930 1.22335281 0.07878923

 Estimated parameter covariance matrix.
            location        scale        shape
location  2.29346581  1.084185707 -0.038081743
scale     1.08418571  1.496592093 -0.008215965
shape    -0.03808174 -0.008215965  0.006207743

 AIC = 1077.656
 BIC = 1086.068

rl = return.level(A, return.period = c(2, 5, 10, 15, 25, 50, 100))
rl
GEV model fitted to  dat_
Data are assumed to be  stationary
[1] "Return Levels for period units in years"
  2-year level   5-year level  10-year level  15-year level
```

47.04059 67.50520 84.15453 94.83606

25-year level 50-year level 100-year level

109.53600 132.10346 158.24321



Model2

ti <- matrix(ncol = 1, nrow = n)

ti[,1] <- seq(1, n, 1)

model2 <- gev.fit(dat_, ydat = ti, mul = 1)

$model

$model[[1]]

[1] 1

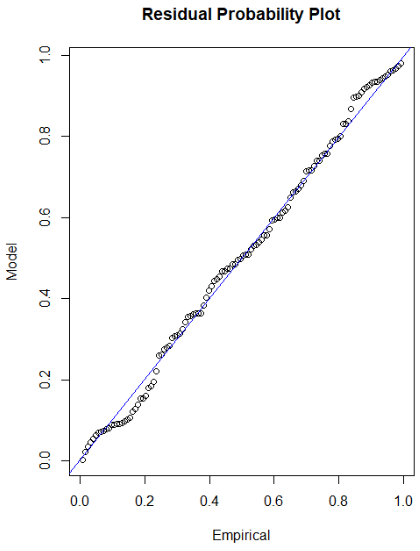
$model[[2]]

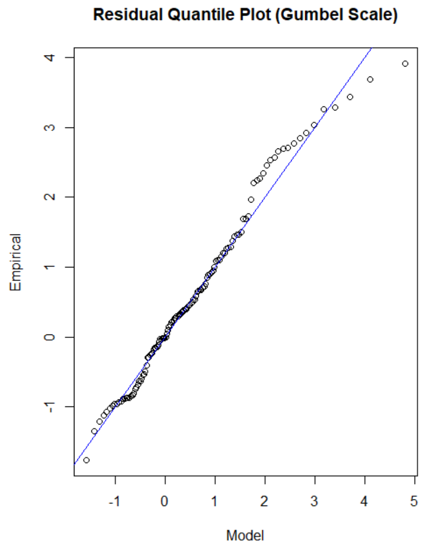
NULL

$model[[3]]

NULL

```
$link
[1] "c(identity, identity, identity)"
$conv
[1] 0
$nllh
[1] 535.0424
$mle
[1] 43.25554673 -0.03410188 14.30071789  0.25511756
$se
[1] 2.04807124 0.02589423 1.24888321 0.08887916

aic.model2 = - 2 * model2$nllh + 2 * 4
aic.model2
[1] -1062.085
bic.model2 = - 2 * model2$nllh + log(length(dat_)) * length(model2$mle)
bic.model2
[1] -1050.869
fit2 <- c(par = model2$mle, se = model2$se, aic = aic.model2, bic = bic.model2)
fit2
        par1          par2          par3          par4
 4.325555e+01 -3.410188e-02  1.430072e+01  2.551176e-01
         se1           se2           se3           se4
 2.048071e+00  2.589423e-02  1.248883e+00  8.887916e-02
         aic           bic
-1.062085e+03 -1.050869e+03

gev.diag(model2)
```





A = fevd(dat_, threshold = NULL, threshold.fun = ~1, location.fun = ~ti,

scale.fun = ~1, shape.fun = ~1, use.phi = FALSE,

type = "GEV", method = "MLE", initial = NULL,

units = NULL, time.units = "weeks", period.basis = "year",

optim.args = NULL, priorFun = NULL,

priorParams = NULL, proposalFun = NULL, proposalParams = NULL,

iter = 9999, weights = 1, blocks = NULL, verbose = FALSE)

A

[1] "Estimation Method used: MLE"

Negative Log-Likelihood Value: 535.0424

Estimated parameters:

mu0	mu1	scale	shape
43.25617224	-0.03415031	14.29981765	0.25519054

Standard Error Estimates:

mu0	mu1	scale	shape
2.04798159	0.02589059	1.24882751	0.08888707

Estimated parameter covariance matrix.

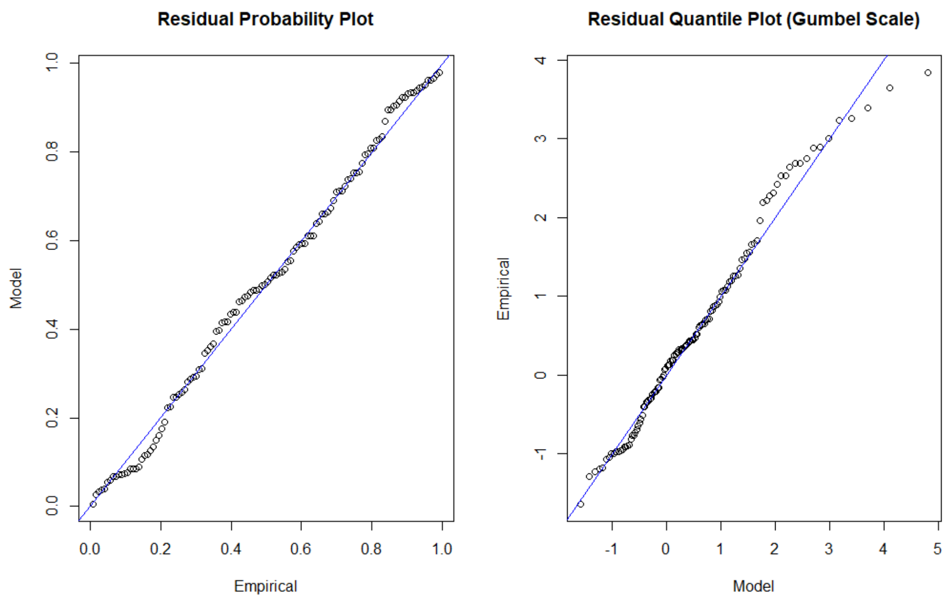
mu0 mu1 scale shape

```
mu0    4.19422860 -0.0360672369  0.950042349 -0.0168481754
mu1   -0.03606724  0.0006703227  0.003881891 -0.0005264805
scale  0.95004235  0.0038818910  1.559570156 -0.0156915268
shape -0.01684818 -0.0005264805 -0.015691527  0.0079009107

 AIC = 1078.085
 BIC = 1089.301

v <- make.qcov(A, vals = list(mu1 = c(-1, 1)))
return.level(A, return.period = c(2, 5, 10, 15, 25, 50, 100), qcov = v)

GEV model fitted to  dat_
Data are assumed to be  non-stationary
[1] "Return Levels for period units in years"
     2-year level 5-year level 10-year level 15-year level
[1,]     48.78431     69.42216      86.76506      98.11888
[2,]     48.71601     69.35386      86.69676      98.05058
     25-year level 50-year level 100-year level
[1,]      114.0058      138.9279       168.5097
[2,]      113.9375      138.8596       168.4414

## Model3 ##
ti = matrix(ncol = 1, nrow = n)
ti[,1] = seq(1,n,1)

ti2 = matrix(ncol = 2, nrow = n)
ti2[,1] = seq(1, n, 1)
ti2[,2] = matrix(ti2[,1] ^ 2)
```

```
model3 <- gev.fit(dat_, ydat = ti2, mul = c(1,2))
$model
$model[[1]]
[1] 1 2
$model[[2]]
NULL
$model[[3]]
NULL

$link
[1] "c(identity, identity, identity)"
$conv
[1] 0
$nllh
[1] 533.7621
$mle
[1] 43.2518339205 -0.0008151201 -0.0004202963 14.0089669640
[5]  0.2750962603
$se
[1] 2.286190e+00 2.719466e-02 4.336015e-05 1.417511e+00 1.213578e-01

aic.model3 = - 2 * model3$nllh + 2 * 5
aic.model3
[1] -1057.524
bic.model3 = - 2 * model3$nllh + log(length(dat_)) * length(model3$mle)
bic.model3
[1] -1043.504
fit3 <- c(par = model3$mle, se = model3$se, aic = aic.model3, bic = bic.model3)
fit3
```

```
      par1          par2          par3          par4
4.325183e+01 -8.151201e-04 -4.202963e-04  1.400897e+01
      par5           se1           se2           se3
2.750963e-01  2.286190e+00  2.719466e-02  4.336015e-05
       se4           se5           aic           bic
1.417511e+00  1.213578e-01 -1.057524e+03 -1.043504e+03
```

gev.diag(model3)

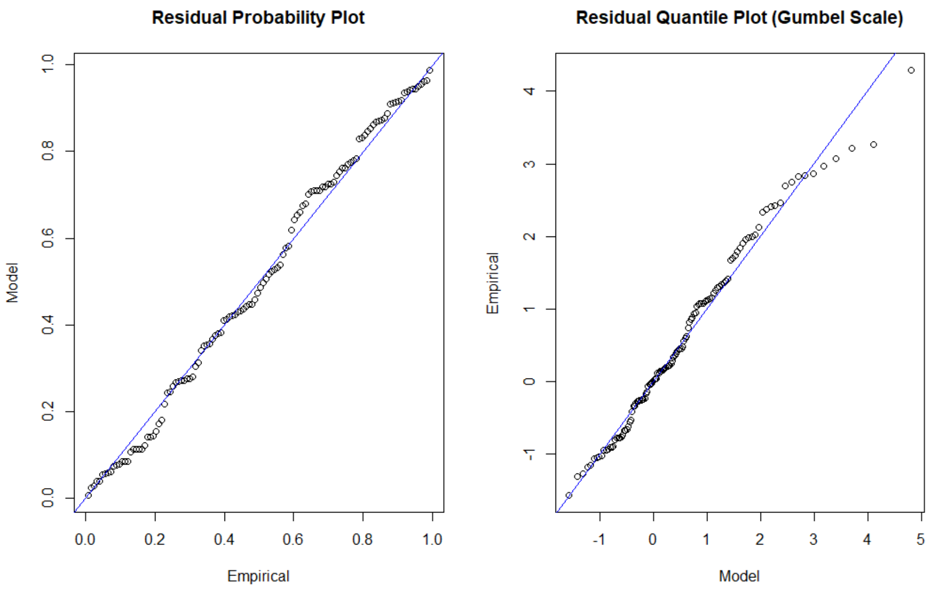



```
A = fevd(dat_, threshold = NULL, threshold.fun = ~1, location.fun = ~ti2[,2],
        scale.fun = ~1, shape.fun = ~1, use.phi = T,
        type = "GEV", method = "MLE", initial = NULL,
        units = NULL, time.units = "weeks", period.basis = "year",
        optim.args = NULL, priorFun = NULL,
        priorParams = NULL, proposalFun = NULL, proposalParams = NULL,
        iter = 9999, weights = 1, blocks = NULL, verbose = FALSE)
A

[1] "Estimation Method used: MLE"
```

```
 Negative Log-Likelihood Value:  533.7393

 Estimated parameters:
          mu0           mu1         scale         shape
43.2300293830 -0.0004282961 13.9989689394  0.2755346344

 Standard Error Estimates:
         mu0          mu1        scale        shape
1.533381e+00 3.862107e-05 1.379224e+00 1.178612e-01

 Estimated parameter covariance matrix.
                mu0           mu1         scale         shape
mu0    2.351256e+00  1.417328e-05  1.413465e+00 -7.784667e-02
mu1    1.417328e-05  1.491587e-09  2.296026e-05 -2.833088e-06
scale  1.413465e+00  2.296026e-05  1.902259e+00 -6.033798e-02
shape -7.784667e-02 -2.833088e-06 -6.033798e-02  1.389126e-02

 AIC = 1075.479
 BIC = 1086.695

v <- make.qcov(A, vals = list(mu1 = c(-1, 1)))
return.level(A, return.period = c(2, 5, 10, 15, 25, 50, 100), qcov = v)

GEV model fitted to  dat_
Data are assumed to be  non-stationary
[1] "Return Levels for period units in years"
     2-year level 5-year level 10-year level 15-year level
[1,]     48.62928     69.23206      86.87476      98.56155
[2,]     48.62842     69.23121      86.87390      98.56069
```

```
     25-year level 50-year level 100-year level
[1,]      115.0736      141.3045       172.8869
[2,]      115.0728      141.3037       172.8861

## Model4 ##
ti = matrix(ncol = 1, nrow = n)
ti[,1] = seq(1, n, 1)

ti2 = matrix(ncol = 2, nrow = n)
ti2[,1] = seq(1, n, 1)
ti2[,2] = (ti2[,1] ^ 2)

model4 <- gev.fit(dat_, ydat = ti, mul = 1, sigl = 1, siglink = exp)
$model
$model[[1]]
[1] 1
$model[[2]]
[1] 1
$model[[3]]
NULL

$link
[1] "c(identity, exp, identity)"
$conv
[1] 0
$nllh
[1] 532.9417
$mle
[1] 38.540272817  0.048393092  2.241066734  0.006388736  0.278334220
```

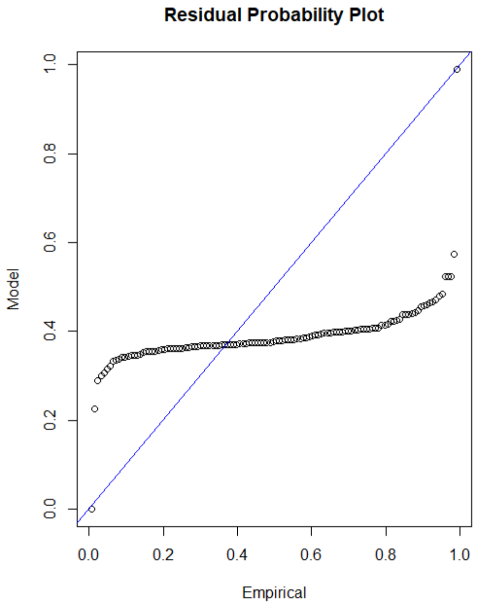
```
$se
[1] 2.571240636 0.046772872 0.215178322 0.002997549 0.102402133

aic.model4 = - 2 * model4$nllh + 2 * 5
aic.model4
[1] -1055.883
bic.model4 = - 2 * model4$nllh + log(length(dat_)) * length(model4$mle)
bic.model4
[1] -1041.863
fit4  <- c(par = model4$mle, se = model4$se, aic = aic.model4, bic = bic.model4)
fit4
         par1          par2          par3          par4
 3.854027e+01  4.839309e-02  2.241067e+00  6.388736e-03
         par5           se1           se2           se3
 2.783342e-01  2.571241e+00  4.677287e-02  2.151783e-01
          se4           se5           aic           bic
 2.997549e-03  1.024021e-01 -1.055883e+03 -1.041863e+03
gev.diag(model4)
```

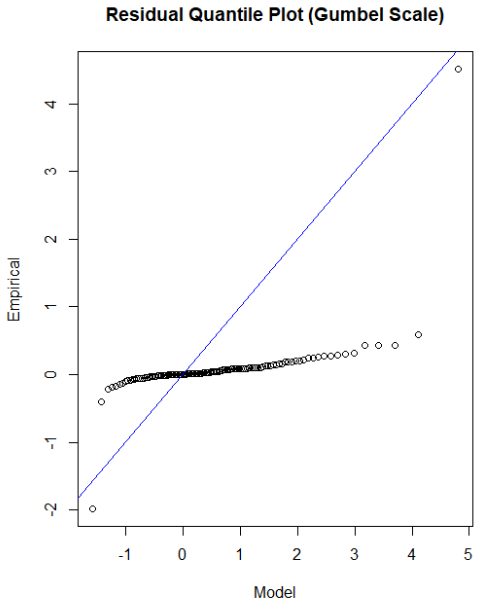


```
A = fevd(dat_, threshold = NULL, threshold.fun = ~1, location.fun = ~ti,
         scale.fun = ~ti, shape.fun = ~1, use.phi = T,
         type = "GEV", method = "MLE", initial = NULL,
         units = NULL, time.units = "weeks", period.basis = "year",
         optim.args = NULL, priorFun = NULL,
         priorParams = NULL, proposalFun = NULL, proposalParams = NULL,
         iter = 9999, weights = 1, blocks = NULL, verbose = FALSE)
A
[1] "Estimation Method used: MLE"

 Negative Log-Likelihood Value:  532.9415

 Estimated parameters:
        mu0          mu1         phi0         phi1        shape
38.526649002  0.048712873  2.239642300  0.006426486  0.279783014



 Standard Error Estimates:
       mu0         mu1        phi0        phi1       shape
2.566159473 0.046774797 0.217222774 0.003072499 0.101560183

 Estimated parameter covariance matrix.
               mu0           mu1          phi0          phi1
mu0    6.585174442 -0.0996951463  0.4516166406 -6.159799e-03
mu1   -0.099695146  0.0021878816 -0.0070634696  1.192108e-04
phi0   0.451616641 -0.0070634696  0.0471857334 -6.068913e-04
phi1  -0.006159799  0.0001192108 -0.0006068913  9.440251e-06
shape -0.082152148  0.0004885915 -0.0064358948  7.442630e-05
               shape
```

```
mu0   -0.0821521476
mu1    0.0004885915
phi0  -0.0064358948
phi1   0.0000744263
shape  0.0103144708
 AIC = 1075.883
 BIC = 1089.903

v <- make.qcov(A, vals = list(mu1 = c(-1, 1)))
return.level(A, return.period = c(2, 5, 10, 15, 25, 50, 100), qcov = v)

GEV model fitted to  dat_
Data are assumed to be  non-stationary
[1] "Return Levels for period units in years"
     2-year level 5-year level 10-year level 15-year level
[1,]     42.12549     56.09116      68.09743      76.07012
[2,]     42.22292     56.18859      68.19485      76.16754
     25-year level 50-year level 100-year level
[1,]      87.35741      105.3355       127.0460
[2,]      87.45484      105.4329       127.1434


 ** ## Model5 ## **
ti = matrix(ncol = 1, nrow = n)
ti[,1] = seq(1, n, 1)

ti2 = matrix(ncol = 2, nrow = n)
ti2[,1] = seq(1, n, 1)
ti2[,2] = (ti2[,1] ^ 2)
```

```
model5 <-  gev.fit(dat_, ydat = ti, sigl = 1, siglink = exp,
                    shl = 1)
$model
$model[[1]]
NULL
$model[[2]]
[1] 1
$model[[3]]
[1] 1

$link
[1] "c(identity, exp, identity)"
$conv
[1] 0

$nllh
[1] 721.0946
$mle
[1] 40.37620829  4.51180451  0.01342888  9.81800716 -0.14601942
$se
[1] 1.414724e-06 2.829257e-06 2.000668e-06 1.414725e-06 2.000622e-06
gev.diag(model5)
```


Residual Probability Plot
Model
Empirical
0.0
0.2
0.4
0.6
0.8
1.0


Residual Quantile Plot (Gumbel Scale)
Empirical
Model
-2
-1
0
1
2
3
4
5

Dindaeng Station PM2.5

การเรียกใช้แพคเกจของโปรแกรม

```
library(extRemes)
library(ismev)
library(aTSA)
library(tseries)
library(trend)
library(e1071)
library(TLMoments)
library(nortest)
library(eva)
library(ADGofTest)
library(lattice)
```

การเรียกใช้ข้อมูล

```
data <- read.csv("C:\\Users\\acer\\Desktop\\Project\\Code\\clean_Dindaeng\\clean_
Dindaeng_pm2.5_week.csv", header = T)
```

การจัดการข้อมูล และ หา summary

```
weather = na.omit(data$PM2.5)
summary(weather)
  Min.   1st Qu.  Median    Mean   3rd Qu.    Max.
  37.00    58.20    73.34     79.03    94.25    202.64
```

การทดสอบเพื่อดูแนวโน้มของข้อมูล

```
mk.test(weather, continuity = T)

Mann-Kendall trend test
```

data: weather

z = -0.8206, n = 312, p-value = 0.4119

alternative hypothesis: true S is not equal to 0

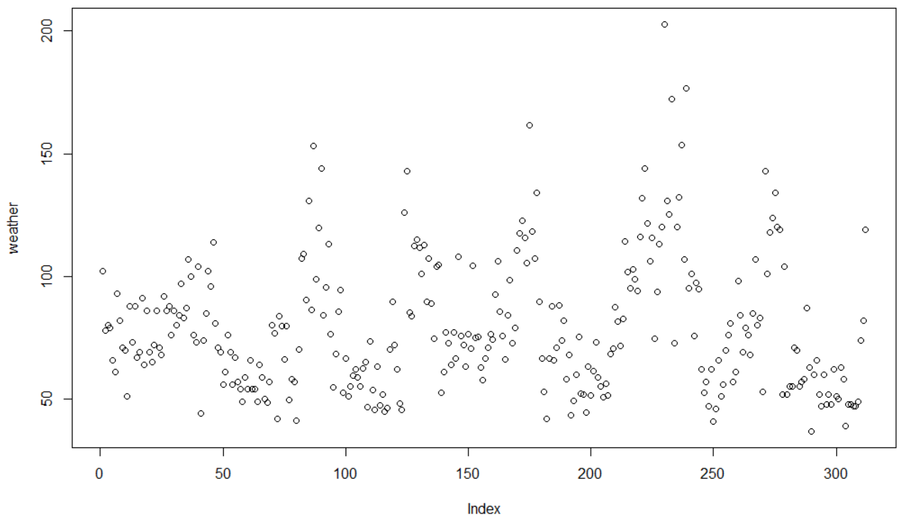
sample estimates:

S varS tau

-1.512000e+03 3.390528e+06 -3.121520e-02

การสร้างกราฟเพื่อดูความหนาแน่นของข้อมูล

plot(weather)



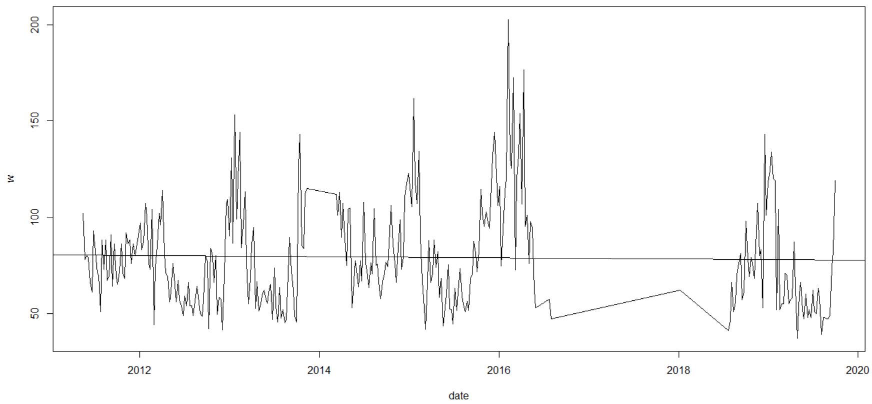
การสร้างกราฟเพื่อดูแนวโน้ม

date <- as.Date(as.character(data$DAY), format = "%Y-%m-%d")

linearmodel <- lm(data$PM2.5 ~ date)

plot(data$PM2.5 ~ date, type = 'l')

abline(linearmodel)



การทดสอบความเป็น Stationary ความเบ้ ความโค้ง

```
stationary.test(weather)
Augmented Dickey-Fuller Test
alternative: stationary

Type 1: no drift no trend
     lag    ADF p.value
[1,]   0 -2.466   0.015
[2,]   1 -1.501   0.143
[3,]   2 -1.153   0.267
[4,]   3 -1.038   0.308
[5,]   4 -0.950   0.339
[6,]   5 -0.843   0.378
Type 2: with drift no trend
     lag   ADF p.value
[1,]   0 -8.56    0.01
[2,]   1 -5.66    0.01
[3,]   2 -4.58    0.01
[4,]   3 -4.27    0.01
[5,]   4 -4.16    0.01
[6,]   5 -3.89    0.01
Type 3: with drift and trend
     lag   ADF p.value
[1,]   0 -8.56  0.0100
[2,]   1 -5.66  0.0100
[3,]   2 -4.58  0.0100
[4,]   3 -4.27  0.0100
[5,]   4 -4.16  0.0100
[6,]   5 -3.89  0.0145
```

Note: in fact, p.value = 0.01 means p.value <= 0.01

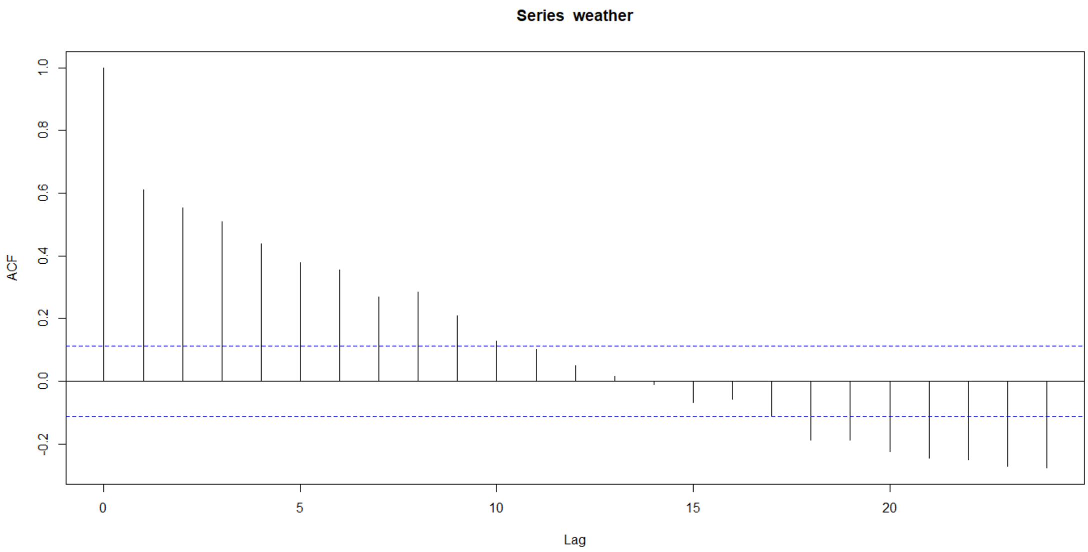
skewness(weather)

[1] 1.135368

kurtosis(weather)

[1] 1.654087

acf(weather)



adf.test(weather)

Augmented Dickey-Fuller Test

data: weather

Dickey-Fuller = -4.0721, Lag order = 6, p-value =

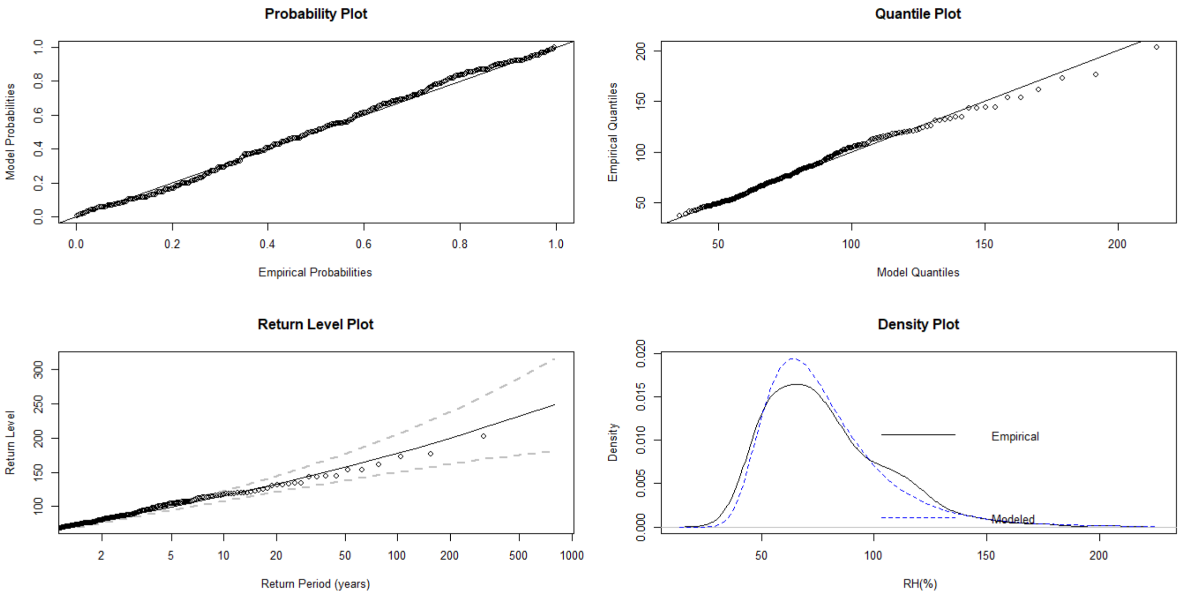
0.01

alternative hypothesis: stationary

การหาค่าประมาณพารามิเตอร์ที่เหมาะสม ระดับการเกิดซ้ำ และ ช่วงความเชื่อมั่น

```
A = gev.fit(weather)
$conv
[1] 0

$nllh
[1] 1430.842
$mle
[1] 65.9725704 19.0423487  0.1022072
$se
[1] 1.25731672 0.96890777 0.05298275

A = fevd(data[,2], threshold = NULL, threshold.fun = ~1, location.fun = ~1,
        scale.fun = ~1, shape.fun = ~1, use.phi = FALSE,
        type = "GEV", method = "MLE", initial = NULL,
        units = NULL, time.units = "months", period.basis = "year",
        optim.args = NULL, priorFun = NULL,
        priorParams = NULL, proposalFun = NULL, proposalParams = NULL,
        iter = 9999, weights = 1, blocks = NULL, verbose = FALSE)

rl = return.level(A, return.period = c(2, 5, 10, 15, 25, 50, 100))
rl
fevd(x = data[, 2], threshold = NULL, threshold.fun = ~1, location.fun = ~1,
    scale.fun = ~1, shape.fun = ~1, use.phi = FALSE, type = "GEV",
    method = "MLE", initial = NULL, units = NULL, time.units = "months",
    period.basis = "year", optim.args = NULL, priorFun = NULL,
    priorParams = NULL, proposalFun = NULL, proposalParams = NULL,
    iter = 9999, weights = 1, blocks = NULL, verbose = FALSE)
    get(paste("return.level.fevd.", newcl, sep = ""))(x = x,
```

```
    return.period = return.period)

 GEV model fitted to  data[, 2]
Data are assumed to be  stationary
[1] "Return Levels for period units in years"
  2-year level   5-year level  10-year level  15-year level
      73.08116       96.83936      114.15474      124.52657
 25-year level  50-year level 100-year level
     138.02168      157.28457      177.82600

loc      = A$results$par[1]
scale  = A$results$par[2]
shape = A$results$par[3]
se.loc      = sqrt(diag(solve(A$results$hessian)))[1]
se.scale  = sqrt(diag(solve(A$results$hessian)))[2]
se.shape = sqrt(diag(solve(A$results$hessian)))[3]

loc.ci      = c(loc - 1.96 * se.loc, loc + 1.96 * se.loc)
loc.ci
location  location
63.50519  68.43357
scale.ci  = c(scale - 1.96 * se.scale, scale + 1.96 * se.scale)
shape.ci = c(shape - 1.96 * se.shape, shape + 1.96 * se.shape)

par(mfrow = c(2,2))
plot(A, "probprob", main = "Probability Plot")
plot(A, "qq", main = "Quantile Plot")
plot(A, "rl", main = "Return Level Plot")
plot(A, "density", xlab = " RH(%)", main = "Density Plot")
```



GOF test

ks.test(data[,2], pgev, loc, scale, shape)

One-sample Kolmogorov-Smirnov test

data: data[, 2]

D = 0.038545, p-value = 0.7429

alternative hypothesis: two-sided

Warning message:

In ks.test(data[, 2], pgev, loc, scale, shape) :

  ties should not be present for the Kolmogorov-Smirnov test

AIC & BIC

A = gev.fit(weather)

aic.A = -2 * A$nllh + 2 * 2

```
bic.A = -2 * A$nllh + log(length(A$data)) * length(A$mle)
fit1  <- c(par = A$mle,
           se  = A$se,
           aic = aic.A,
           bic = bic.A)
fit1
    par1          par2          par3          se1
 6.597257e+01  1.904235e+01  1.022072e-01  1.257317e+00
      se2           se3            aic           bic
 9.689078e-01  5.298275e-02  -2.857684e+03  -2.844455e+03
aic.A
[1] -2857.684
bic.A
[1] -2844.455
```

Pathumwan PM2.5

การเรียกใช้แพคเกจของโปรแกรม

```
library(extRemes)
library(ismev)
library(aTSA)
library(tseries)
library(trend)
library(e1071)
library(TLMoments)
library(nortest)
library(eva)
library(ADGofTest)
library(lattice)
```

การเรียกใช้ข้อมูล

```
data <-
read.csv("C:\\Users\\acer\\Desktop\\Project\\Code\\clean_Jula_Pathumwan\\clean_
Jula_Pathumwan_pm2.5_week.csv", header = T)
```

การจัดการข้อมูล และ หา summary

```
weather = na.omit(data$PM2.5)
summary(weather)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  28.00   41.00   52.00   57.79   68.50  130.00 
```

การทดสอบเพื่อดูแนวโน้มของข้อมูล

```
mk.test(weather, continuity = T)

Mann-Kendall trend test

data:  weather
z = -0.83033, n = 122, p-value = 0.4064
```

alternative hypothesis: true S is not equal to 0

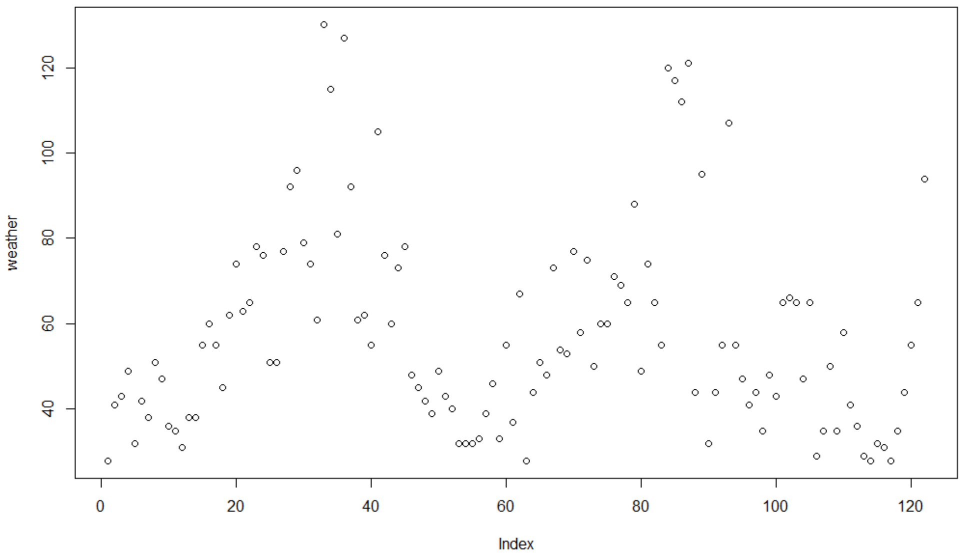
sample estimates:

S varS tau

-3.760000e+02 2.039687e+05 -5.142816e-02

การสร้างกราฟเพื่อดูความหนาแน่นของข้อมูล

plot(weather)



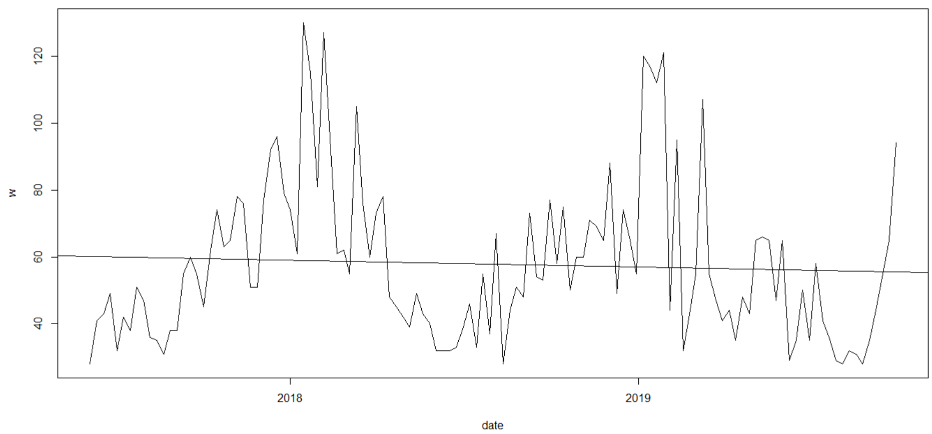
การสร้างกราฟเพื่อดูแนวโน้ม

date <- as.Date(as.character(data$DAY), format = "%Y-%m-%d")

linearmodel <- lm(data$PM2.5 ~ date)

plot(data$PM2.5 ~ date, type = 'l')

abline(linearmodel)



การทดสอบความเป็น Stationary ความเบ้ ความโด้ง

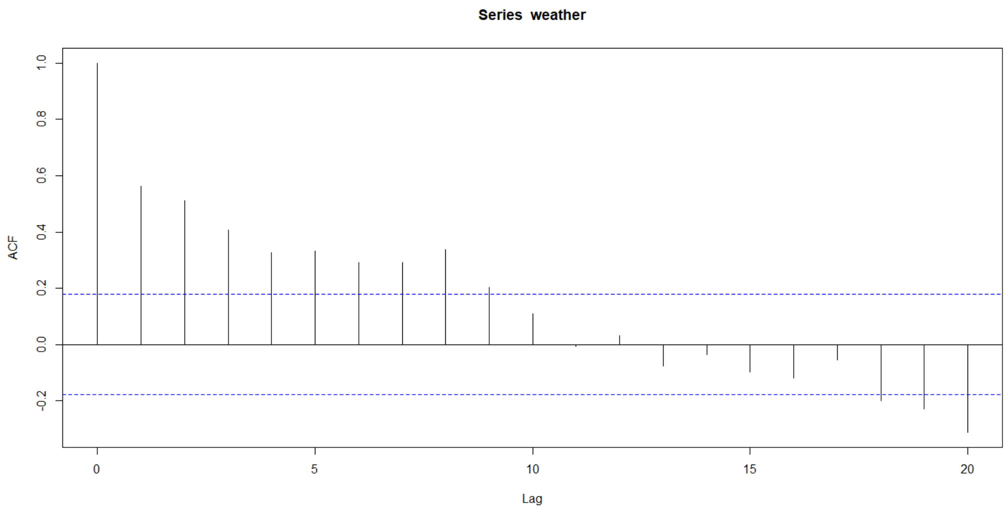
```
stationary.test(weather)
Augmented Dickey-Fuller Test
alternative: stationary

Type 1: no drift no trend
     lag    ADF p.value
[1,]   0 -1.668  0.0917
[2,]   1 -0.897  0.3572
[3,]   2 -0.720  0.4210
[4,]   3 -0.665  0.4404
[5,]   4 -0.467  0.5094
Type 2: with drift no trend
     lag   ADF p.value
[1,]   0 -5.64  0.0100
[2,]   1 -3.69  0.0100
[3,]   2 -3.26  0.0209
[4,]   3 -3.09  0.0318
[5,]   4 -2.74  0.0748
Type 3: with drift and trend
     lag   ADF p.value
[1,]   0 -5.64  0.0100
[2,]   1 -3.69  0.0282
[3,]   2 -3.26  0.0817
[4,]   3 -3.09  0.1231
[5,]   4 -2.77  0.2529
----
Note: in fact, p.value = 0.01 means p.value <= 0.01
skewness(weather)
```

[1] 1.120077

kurtosis(weather)

[1] 0.8180595

acf(weather)



adf.test(weather)

Augmented Dickey-Fuller Test

data: weather

Dickey-Fuller = -2.7743, Lag order = 4, p-value = 0.2549

alternative hypothesis: stationary

การสร้างโมเดล หาค่าประมาณพารามิเตอร์ที่เหมาะสม AIC&BIC และ ระดับการเกิดซ้ำ

dat <-

read.csv("C:\\Users\\acer\\Desktop\\Project\\Code\\clean_Jula_Pathumwan\\clean

_Jula_Pathumwan_pm2.5_week.csv", header = T)

length(dat)

n = nrow(dat)

```
dat_ = dat[,2]

## Model1 ##
model1 <- gev.fit(dat_)
$conv
[1] 0
$nllh
[1] 539.9177
$mle
[1] 45.7740743 15.5142243  0.1824851
$se
[1] 1.67278013 1.34146288 0.09512303

aic.model1 = - 2 * model1$nllh + 2 * 3
aic.model1
[1] -1073.835
bic.model1 = - 2 * model1$nllh + log(length(dat_)) * length(model1$mle)
bic.model1
[1] -1065.423
fit1  <- c(par = model1$mle, se = model1$se, aic = aic.model1, bic = bic.model1)
fit1
         par1          par2          par3           se1           se2 
 4.577407e+01  1.551422e+01  1.824851e-01  1.672780e+00  1.341463e+00 
          se3           aic           bic 
 9.512303e-02 -1.073835e+03 -1.065423e+03
```

การหาความเหมาะสมของตัวแบบ

```
### GOF TEST ###
ks.test(dat$PM2.5, pgev, model1$mle[1], model1$mle[2], model1$mle[3])

        One-sample Kolmogorov-Smirnov test

data:  dat$PM2.5
D = 0.053233, p-value = 0.8798
alternative hypothesis: two-sided

Warning message:
In ks.test(dat$PM2.5, pgev, model1$mle[1], model1$mle[2], model1$mle[3]) :
 ties should not be present for the Kolmogorov-Smirnov test
A = fevd(dat_, threshold = NULL, threshold.fun = ~1, location.fun = ~1,
        scale.fun = ~1, shape.fun = ~1, use.phi = FALSE,
        type = "GEV", method = "MLE", initial = NULL,
        units = NULL, time.units = "weeks", period.basis = "year",
        optim.args = NULL, priorFun = NULL,
        priorParams = NULL, proposalFun = NULL, proposalParams = NULL,
        iter = 9999, weights = 1, blocks = NULL, verbose = FALSE)
A
[1] "Estimation Method used: MLE"

 Negative Log-Likelihood Value:  539.9177

 Estimated parameters:
  location       scale       shape
45.7719006 15.5092578  0.1825668
```

Standard Error Estimates:

location scale shape

1.67258508 1.34067798 0.09509403

Estimated parameter covariance matrix.

	location	scale	shape
location	2.79754084	1.34622403	-0.067210652
scale	1.34622403	1.79741744	-0.035766391
shape	-0.06721065	-0.03576639	0.009042875

AIC = 1085.835

BIC = 1094.247

rl = return.level(A, return.period = c(2, 5, 10, 15, 25, 50, 100))

rl

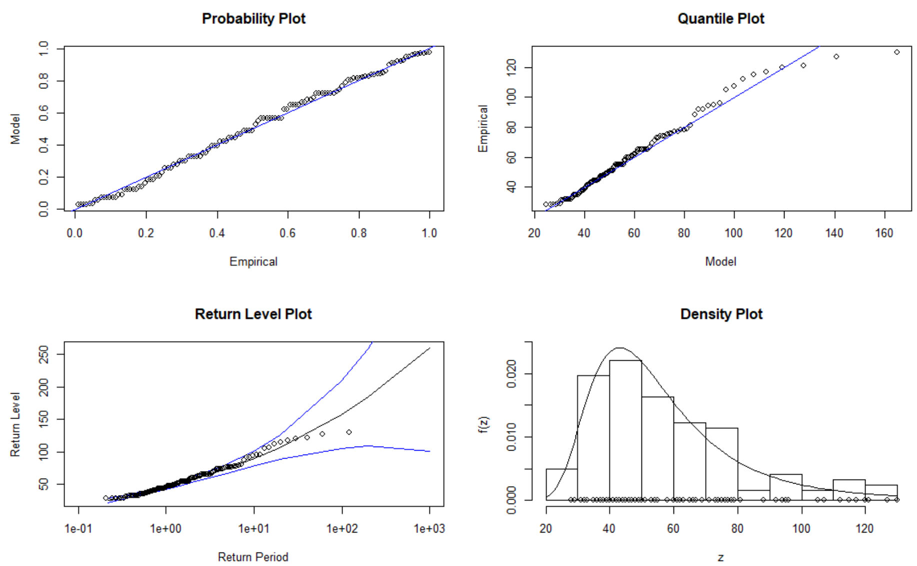
GEV model fitted to dat_

Data are assumed to be stationary

[1] "Return Levels for period units in years"

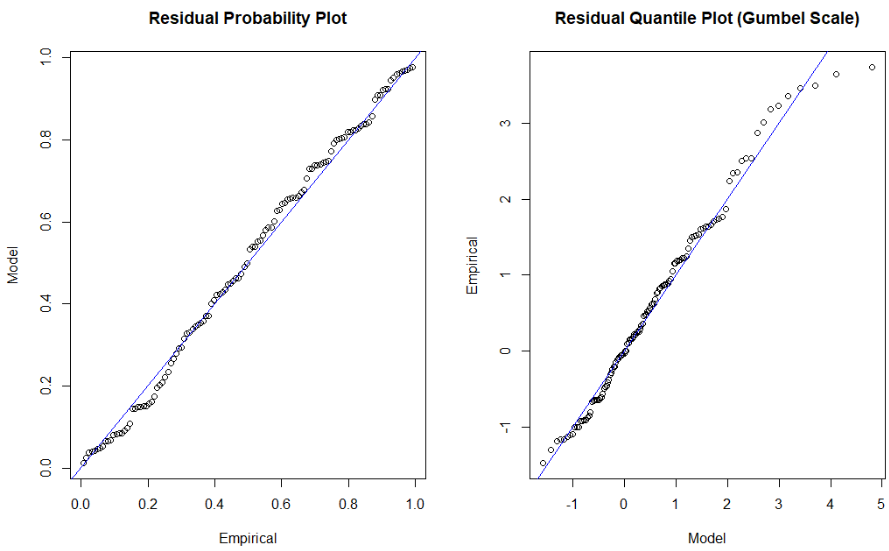
2-year level	5-year level	10-year level	15-year level
51.65074	72.53180	88.93452	99.22998

25-year level	50-year level	100-year level
113.14669	134.01968	157.56623



```
## Model2 ##
ti <- matrix(ncol = 1, nrow = n)
ti[,1] <- seq(1, n, 1)
model2 <- gev.fit(dat_, ydat = ti, mul = 1)
$model
$model[[1]]
[1] 1
$model[[2]]
NULL
$model[[3]]
NULL

$link
[1] "c(identity, identity, identity)"
$conv
[1] 0

$nllh
[1] 539.3489
$mle
[1] 47.73860672 -0.03215919 15.38613693  0.18859061
$se
[1] 2.52962055 0.03002748 1.33510760 0.09551639

aic.model2 = - 2 * model2$nllh + 2 * 4
aic.model2
[1] -1070.698
bic.model2 = - 2 * model2$nllh + log(length(dat_)) * length(model2$mle)
bic.model2
```

```
[1] -1059.482
fit2 <- c(par = model2$mle, se = model2$se, aic = aic.model2, bic = bic.model2)
fit2
        par1          par2          par3          par4           se1
 4.773861e+01 -3.215919e-02  1.538614e+01  1.885906e-01  2.529621e+00
          se2           se3           se4           aic           bic
 3.002748e-02  1.335108e+00  9.551639e-02 -1.070698e+03 -1.059482e+03

gev.diag(model2)
```




```
A = fevd(dat_, threshold = NULL, threshold.fun = ~1, location.fun = ~ti,
         scale.fun = ~1, shape.fun = ~1, use.phi = FALSE,
         type = "GEV", method = "MLE", initial = NULL,
         units = NULL, time.units = "weeks", period.basis = "year",
         optim.args = NULL, priorFun = NULL,
         priorParams = NULL, proposalFun = NULL, proposalParams = NULL,
         iter = 9999, weights = 1, blocks = NULL, verbose = FALSE)
```

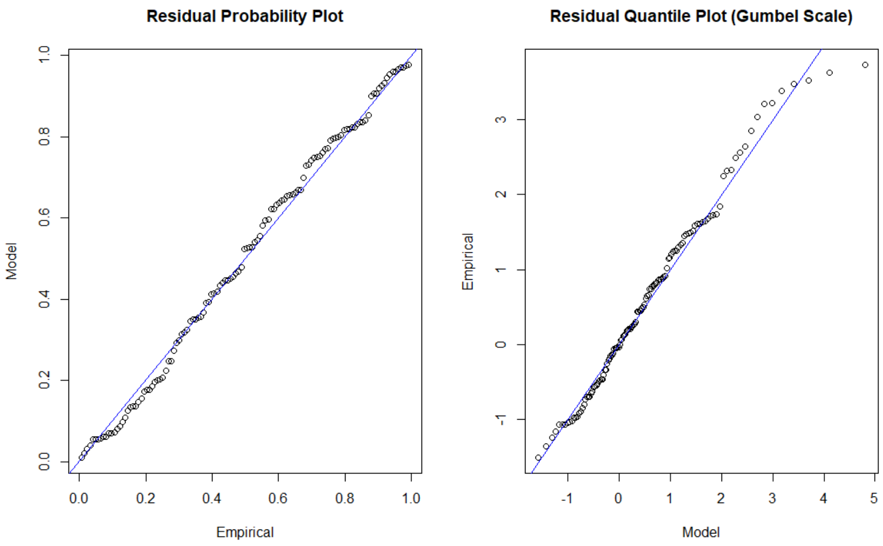
```
A

[1] "Estimation Method used: MLE"

 Negative Log-Likelihood Value:  539.3489

 Estimated parameters:
        mu0         mu1       scale       shape 
47.73905340 -0.03219232 15.38551533  0.18861465 

 Standard Error Estimates:
       mu0        mu1      scale      shape 
2.52972942 0.03002837 1.33507565 0.09551820 

 Estimated parameter covariance matrix.
              mu0           mu1        scale         shape
mu0    6.39953094 -0.0573365181  1.438266599 -0.0788726912
mu1   -0.05733652  0.0009017028 -0.001576143  0.0001927387
scale  1.43826660 -0.0015761431  1.782426986 -0.0353319235
shape -0.07887269  0.0001927387 -0.035331924  0.0091237272

 AIC = 1086.698 
 BIC = 1097.914 

v <- make.qcov(A, vals = list(mu1 = c(-1, 1)))
return.level(A, return.period = c(2, 5, 10, 15, 25, 50, 100), qcov = v)
GEV model fitted to  dat_ 
Data are assumed to be  non-stationary 
[1] "Return Levels for period units in years"
```

```
     2-year level 5-year level 10-year level 15-year level
[1,]     53.60972     74.44392     90.90222     101.2689
[2,]     53.54533     74.37954     90.83784     101.2045
     25-year level 50-year level 100-year level
[1,]      115.3223      136.4791      160.4472
[2,]      115.2579      136.4147      160.3828
```

Model3

```
ti = matrix(ncol = 1, nrow = n)
ti[,1] = seq(1, n, 1)

ti2 = matrix(ncol = 2, nrow = n)
ti2[,1] = seq(1, n, 1)
ti2[,2] = matrix(ti2[,1] ^ 2)
model3 <- gev.fit(dat_, ydat = ti2, mul = c(1,2))
$model
$model[[1]]
[1] 1 2
$model[[2]]
NULL
$model[[3]]
NULL
$link
[1] "c(identity, identity, identity)"
$conv
[1] 0
$nllh
[1] 538.4487
$mle
```

```
[1] 47.9548105670 -0.0004287688 -0.0004157332 15.2712546887 0.1880522342
$se
[1] 1.1248037       NaN       NaN 0.7130063       NaN

aic.model3 = - 2 * model3$nllh + 2 * 5
aic.model3
[1] -1066.897
bic.model3 = - 2 * model3$nllh + log(length(dat_)) * length(model3$mle)
bic.model3
[1] -1052.877
fit3 <- c(par = model3$mle, se = model3$se, aic = aic.model3, bic = bic.model3)
fit3
        par1          par2          par3          par4          par5
 4.795481e+01 -4.287688e-04 -4.157332e-04  1.527125e+01  1.880522e-01
         se1           se2           se3           se4           se5
 1.124804e+00           NaN           NaN  7.130063e-01           NaN
         aic           bic
-1.066897e+03 -1.052877e+03

gev.diag(model3)
```



A = fevd(dat_, threshold = NULL, threshold.fun = ~1, location.fun = ~ti2[,2],

scale.fun = ~1, shape.fun = ~1, use.phi = T,

type = "GEV", method = "MLE", initial = NULL,

units = NULL, time.units = "weeks", period.basis = "year",

optim.args = NULL, priorFun = NULL,

priorParams = NULL, proposalFun = NULL, proposalParams = NULL,

iter = 9999, weights = 1, blocks = NULL, verbose = FALSE)

A

[1] "Estimation Method used: MLE"

Negative Log-Likelihood Value: 538.4401

Estimated parameters:

```
         mu0           mu1         scale         shape 
47.9643378273 -0.0004213954 15.2758773275  0.1879371533 
```

Standard Error Estimates:

```
          mu0           mu1         scale         shape
1.8725211755 0.0002235233 1.3654291534 0.0985964454

 Estimated parameter covariance matrix.
                  mu0            mu1          scale          shape
mu0    3.5063355529 -2.016955e-04  9.716510e-01 -3.574282e-02
mu1   -0.0002016955  4.996265e-08  8.058672e-05 -6.888410e-06
scale  0.9716509888  8.058672e-05  1.864397e+00 -4.355168e-02
shape -0.0357428239 -6.888410e-06 -4.355168e-02  9.721259e-03

 AIC = 1084.88
 BIC = 1096.096
v <- make.qcov(A, vals = list(mu1 = c(-1, 1)))
return.level(A, return.period = c(2, 5, 10, 15, 25, 50, 100), qcov = v)
GEV model fitted to  dat_
Data are assumed to be  non-stationary
[1] "Return Levels for period units in years"
     2-year level 5-year level 10-year level 15-year level
[1,]     53.76090     74.43328      90.75346      101.0291
[2,]     53.76005     74.43244      90.75261      101.0283
     25-year level 50-year level 100-year level
[1,]      114.9546      135.9101       159.6388
[2,]      114.9537      135.9092       159.6380
```

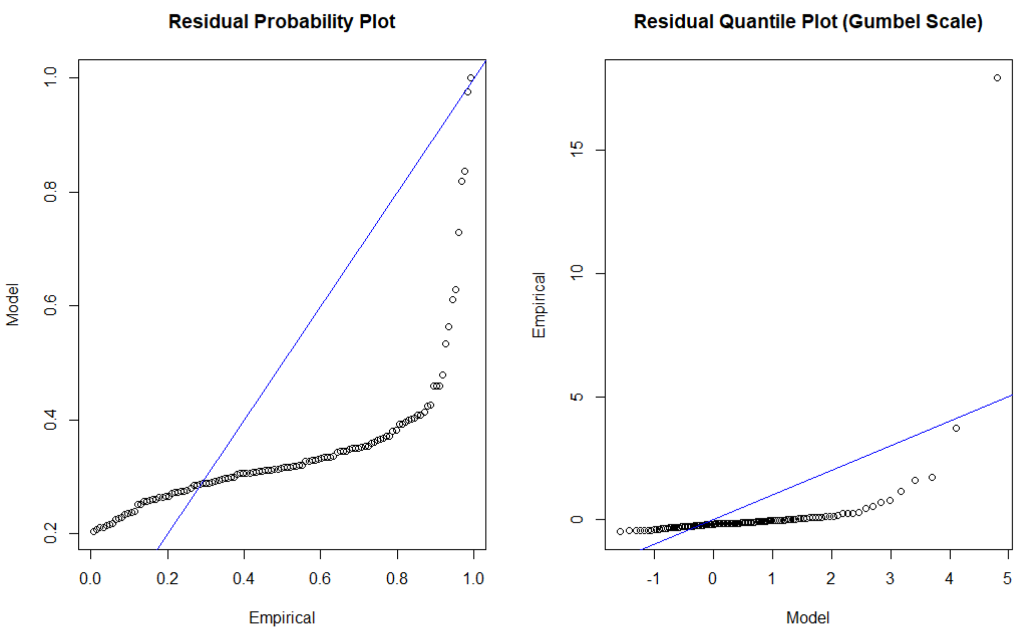
Model4

```
ti = matrix(ncol = 1, nrow = n)
ti[,1] = seq(1, n, 1)
ti2 = matrix(ncol = 2, nrow = n)
ti2[,1] = seq(1, n, 1)
```

```
ti2[,2] = (ti2[,1] ^ 2)

model4 <- gev.fit(dat_, ydat = ti, mul = 1, sigl = 1, siglink = exp)
$model
$model[[1]]
[1] 1
$model[[2]]
[1] 1
$model[[3]]
NULL

$link
[1] "c(identity, exp, identity)"

$conv
[1] 10



Warning message:
In sqrt(diag(z$cov)) : NaNs produced
aic.model4 = - 2 * model4$nllh + 2 * 5
aic.model4
[1] -1374.823
bic.model4 = - 2 * model4$nllh + log(length(dat_)) * length(model4$mle)
bic.model4
[1] -1360.803
fit4  <- c(par = model4$mle, se = model4$se, aic = aic.model4, bic = bic.model4)
fit4
```

```
     par1          par2          par3          par4          par5
 5.934964e+01  2.058760e-01  5.105491e+00 -7.016534e-03 -2.047888e+00
      se1           se2           se3           se4           se5
 5.397559e-04           NaN  5.003058e-04  2.000660e-06  2.000679e-06
      aic           bic
-1.374823e+03 -1.360803e+03
```

gev.diag(model4)



Model5

ti = matrix(ncol = 1, nrow = n)

ti[,1] = seq(1, n, 1)

ti2 = matrix(ncol = 2, nrow = n)

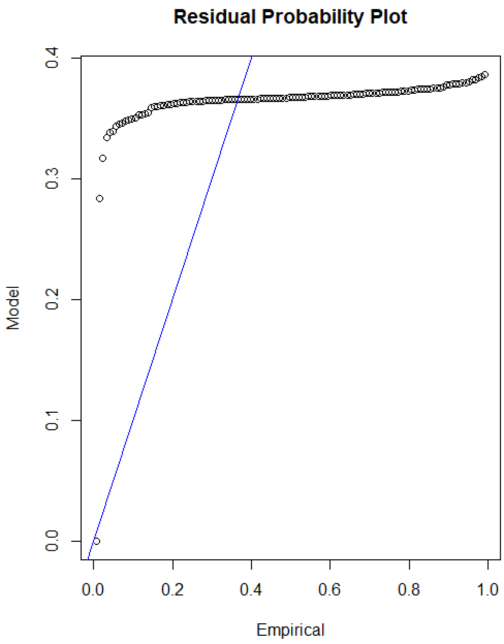
ti2[,1] = seq(1, n, 1)

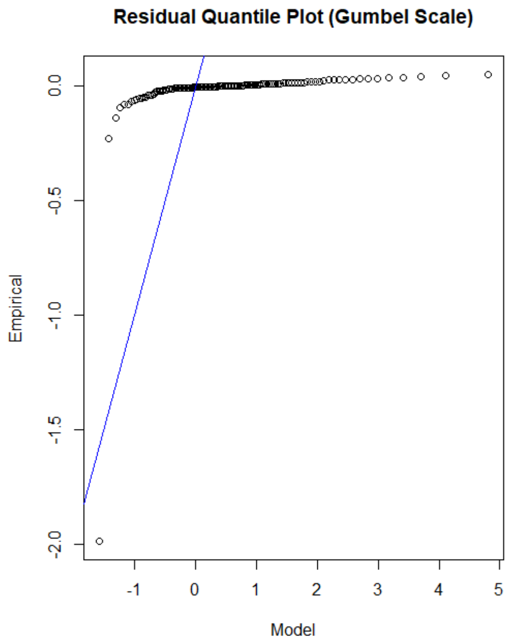
```
ti2[,2] = (ti2[,1] ^ 2)
model5 <-  gev.fit(dat_, ydat = ti, sigl = 1, siglink = exp, shl = 1)
$model
$model[[1]]
NULL
$model[[2]]
[1] 1
$model[[3]]
[1] 1


$link
[1] "c(identity, exp, identity)"

$conv
[1] 10

aic.model5 = - 2 * model5$nllh + 2 * 5
aic.model5
[1] -1878.224
bic.model5 = - 2 * model5$nllh + log(length(dat_)) * length(model5$mle)
bic.model5
[1] -1864.204
Fit5  <- c(par = model5$mle, se = model5$se, aic = aic.model5, bic = bic.model5)
Fit5
        par1          par2          par3          par4          par5
 5.231766e+01  5.710866e+00  2.222222e-02  1.126063e+01  6.036289e-01
         se1           se2           se3           se4           se5
 2.000922e-06  2.000921e-06  2.000899e-06  2.000925e-06  2.000922e-06
```

aic bic

-1.878224e+03 -1.864204e+03

gev.diag(model5)





ภาคผนวก ข

โปสเตอร์แสดงผลงานวิชาการ

โปสเตอร์เข้าร่วมการนำเสนอผลงานวิจัยประเภท Poster Presentation ในโครงการ “Science Exhibition Day 2019” วันที่ 6 กุมภาพันธ์ พ.ศ. 2563 ณ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ