



ผลที่ได้จากการจำแนกประเภทข้อมูลด้วยคอมพิวเตอร์

(Results Obtained from Classification by Data Processing)

ในบทที่ 3 เราได้กล่าวถึงขั้นตอนในการวิเคราะห์ข้อมูลด้วยวิธีการต่าง ๆ ตลอดจน การจำแนกตะกอนแขวนลอยบริเวณอ่างเก็บน้ำเขื่อนภูมิพล โดยใช้ข้อมูล ซี.ซี.ที. วิธีการที่กล่าวถึง ประกอบด้วย

- การเขียน Decision Tree Algorithm
- เทคนิคของ Maximum Likelihood Ratio
- การทำ Multi-channel Density Slicing

ซึ่งผลการจำแนกได้แสดงด้วยแผนภาพในรูปที่ 29 ถึงรูปที่ 34

สำหรับในบทนี้เราจะได้กล่าวถึงการนำข้อมูลจากรายงานการสำรวจตะกอน อ่างเก็บน้ำเขื่อนภูมิพลของการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย มาทำการ เปรียบเทียบและ หาความสัมพันธ์ระหว่างกัน กับผลที่ได้จากการจำแนกประเภทข้อมูลในบทที่ 3 ซึ่งข้อมูล ซี.ซี.ที. ที่ใช้เป็นข้อมูลที่ได้รับการบันทึกโดยดาวเทียมมา ในช่วงเวลาที่ กฟผ. ดำเนินการสำรวจตะกอนพอดี

4.1 ข้อมูลจากการสำรวจตะกอนของ กฟผ.

ตารางที่ 9 เป็นการสรุปผลจากการสำรวจตะกอนฯ ซึ่งประกอบด้วยขนาด พื้นที่ของบริเวณสำรวจ จำนวนจุดที่ทำการเก็บตัวอย่างน้ำมาวิเคราะห์ และค่าเฉลี่ย ความเข้มข้นของปริมาณตะกอนในบริเวณจุดสำรวจนั้น ตลอดจนช่วงเวลาทำการสำรวจ โดยแสดงเฉพาะบริเวณสำรวจที่ 1 ถึง 6 และ 11, 12 ซึ่งการจำแนกประเภทข้อมูลของเราครอบคลุมไปถึงเท่านั้น

บริเวณ สำรวจ	ขนาดพื้นที่ (กม.) ²	จำนวนจุดที่ ทำการเก็บตัวอย่าง	ค่าเฉลี่ยความเข้มข้น ของตะกอน (ppm)	เดือนที่ทำการ สำรวจ
1	1.26	10	27	ธ.ค. 2519
2	1.50	14	55	ม.ค. 2520
3	0.30	12	43	ธ.ค. 2519
4	0.48	11	15	ม.ค. 2520
5	1.12	10	74	"
6	0.30	10	284	"
11	0.25	9	134	เม.ย. 2520
12	0.97	18	28	พ.ค. 2520

ตารางที่ 9 สรุปข้อมูลจากการสำรวจตะกอนฯ ของ กฟผ.

4.2 การกำหนดบริเวณจุดสำรวจต่าง ๆ ลงบนภาพจำแนกประเภทข้อมูลเพื่อทำการเปรียบเทียบ

การวิเคราะห์ข้อมูล ซี.ซี.ที. ในครั้งนี้ ยังไม่มีโอกาสในการทำจุดอ้างอิงภาคพื้นดิน (ground reference point) เพื่อทำ geometrical correction ของข้อมูล ซี.ซี.ที. เพื่อใช้ในการเปรียบเทียบตำแหน่งที่แน่นอนระหว่างข้อมูล ซี.ซี.ที. กับข้อมูลภาคพื้นดินอื่น ๆ นอกจากนั้น การใช้เครื่องพิมพ์คอมพิวเตอร์ (line-printer) ในการพิมพ์ภาพจึงต้องพิมพ์ด้วยการกำหนดตัวอักษรหรือสัญลักษณ์แทนแต่ละจุดภาพของข้อมูลนั้น ๆ ปัญหาขนาดของตัวอักษรที่ไม่เป็นสัดส่วนเดียวกันกับขนาดพื้นที่ของแต่ละจุดภาพ การเว้นช่องไฟ ตลอดจนการขึ้นบรรทัดใหม่ของเครื่องพิมพ์ ยังผลให้เกิดความเพี้ยน (distortion) ของภาพ ทำให้รูปลักษณะของภาพผิดสเกลไปบ้าง

การกำหนดขอบเขตและตำแหน่งของบริเวณสำรวจต่าง ๆ ในวิทยานิพนธ์ได้ อาศัยข้อมูลเท่าที่มีอยู่ คือ

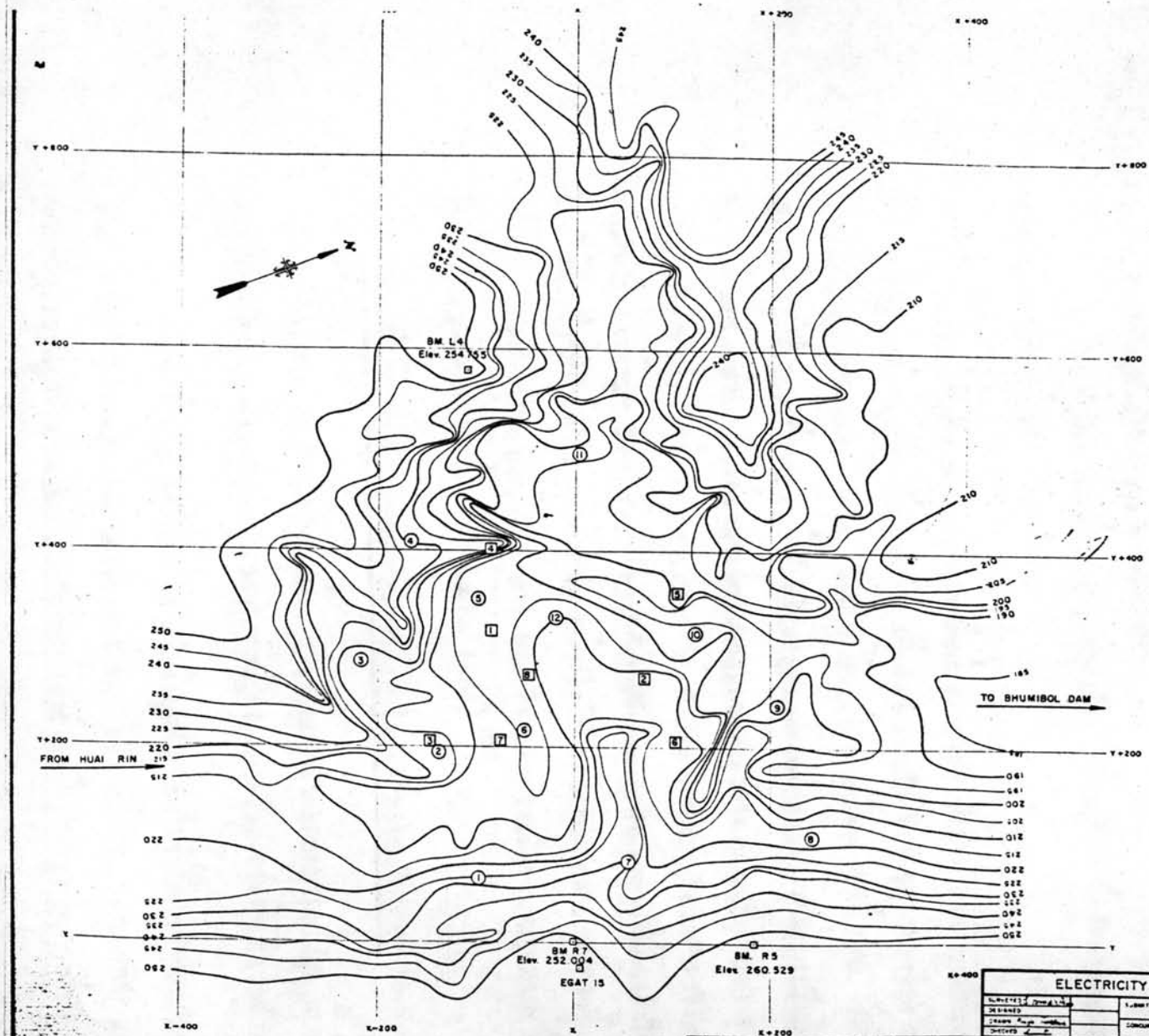
- ก. แผนผังแสดงจุดสำรวจ (ดูรูปที่ 2)
- ข. แผนผังบริเวณจุดสำรวจ (site location map) ซึ่งแสดงแผนที่ห้องอ่างพร้อมกับจุดที่ทำการสูบน้ำ
- ค. แผนที่ภูมิประเทศที่ครอบคลุมอ่างเก็บน้ำเขื่อนภูมิพล ซึ่งจัดทำโดยกรมแผนที่ทหาร ขนาดมาตราส่วน 1 : 50,000

เราไม่สามารถกำหนดขอบเขตที่แน่นอนของบริเวณสำรวจลงในภาพพิมพ์จำแนกประเภทข้อมูล อย่างถูกต้องเพียงตรงได้ เนื่องจากแผนผังบริเวณจุดสำรวจของ กฟผ. เป็นแผนที่แบบ arbitrary grid โดยมีจุดอ้างอิงเป็นหมุดคอนกรีตที่ กฟผ. จัดทำขึ้นเอง (ดูตัวอย่างรูปที่ 35) ซึ่งหมุดเหล่านั้นมิได้ระบุตำแหน่งกริดสากลไว้สำหรับเปรียบเทียบ กับแผนที่ภูมิประเทศ ทำให้ไม่สามารถกำหนดบริเวณสำรวจลงในแผนที่ภูมิประเทศ เพื่อเปรียบเทียบกลับไปยังภาพถ่ายดาวเทียมได้ ดังนั้นการกำหนดบริเวณสำรวจในภาพจำแนกประเภทข้อมูล จึงทำได้แต่เพียงอาศัยรูปลักษณะและค่าลวดลายที่อ้างอิงถึงในรายงาน โดยใช้แผนที่ภูมิประเทศประกอบเท่านั้น อย่างไรก็ตาม เราสามารถกำหนดขนาดพื้นที่ได้อย่างคร่าว ๆ โดยคำนวณขนาดกว้าง - ยาว ของบริเวณสำรวจเป็นจำนวนจุดภาพ ซึ่งในที่นี้ ถือว่าแต่ละจุดภาพแทนพื้นที่ขนาด 56 x 79 เมตร (17)

4.3 ข้อมูลที่ได้จากการจำแนกตะกอนแขวนลอยโดยเครื่องคอมพิวเตอร์

เมื่อกำหนดบริเวณสำรวจต่าง ๆ ลงบนภาพจำแนกประเภทข้อมูลได้แล้ว เราสามารถคำนวณเปอร์เซ็นต์ของจำนวนจุดภาพที่ถูกจำแนกเป็นตะกอนต่อบริเวณสำรวจนั้น ๆ ได้

ตารางที่ 10, 11 และ 12 เป็นข้อมูลที่คำนวณได้จากภาพจำแนกประเภทข้อมูล โดยการเขียน decision tree, M.L.R. และ M.D.S. ตามลำดับ ประกอบด้วยจำนวนจุดภาพที่ถูกจำแนกเป็นน้ำและตะกอน เปอร์เซ็นต์ของจำนวนจุดภาพที่เป็นตะกอนต่อ



LEGEND

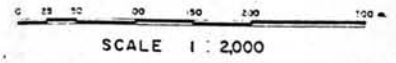
- HORIZONTAL CONTROL POINT WITH ELEVATION
- ↑ APPROXIMATE MAGNETIC NORTH
- ① POINT OF BED LOAD SAMPLING NUMBER 1
- ② POINT OF SUSPENDED SAMPLING NUMBER 1

NOTES

1. HORIZONTAL CONTROL POINT LOCATIONS-
ARBITRARY GRID

	X	Y	ELEV-m
BM R7	0.00	0.00	252.004
BM R5	+184.91	0.00	260.529
BM L4	-109.30	+582.25	254.755

2. ELEVATIONS SHOWN IN METERS ABOVE MEAN-
SEA LEVEL, BY REFERENCE TO EGAT
RESERVOIR LEVEL GAGE
3. FIELD OBSERVATIONS MADE IN DECEMBER 1976
4. MAXIMUM WATER SURFACE ELEVATION
+260.00 M (M.S.L.)



ELECTRICITY GENERATING AUTHORITY OF THAILAND			
<small>REVISIONS</small>	<small>APPROVED</small>	<small>DATE</small>	<small>BY</small>
BHUMBOL RESERVOIR SEDIMENT SURVEY		SITE 3	
OUTLET OF HUAI RIN			
<small>REPLACING</small>		<small>DATE</small>	
<small>DWG. No. HD-88-003</small>			

รูปที่ 35 ตัวอย่างแผนที่บริเวณจุดสำรวจที่ 3 ของ ก.ผ.

จำนวนจุดภาพทั้งหมดของแต่ละบริเวณสำรวจ

บริเวณ สำรวจ	จำแนกเป็น		% ของตะกอน ที่พื้นที่	จำนวนจุดภาพ ทั้งหมด	คิดเป็นขนาด พื้นที่ (กม.) ²
	น้ำ	ตะกอน			
1	183	105	36	288	1.27
2	293	57	16	350	1.55
3	48	21	30	69	0.30
4	102	10	9	112	0.49
5	227	32	12	259	1.15
6	15	48	76	63	0.28
11	33	19	36	52	0.26
12	167	37	18	204	0.90

ตารางที่ 10 ข้อมูลจากภาพจำแนกประเภทโดยวิธีเขียน decision tree

บริเวณ สำรวจ	จำแนกเป็น		% ของตะกอน ที่พื้นที่	จำนวนจุดภาพ ทั้งหมด	คิดเป็นขนาด พื้นที่ (กม.) ²
	น้ำ	ตะกอน			
1	209	74	26	283	1.25
2	315	44	12	359	1.59
3	55	14	20	69	0.30
4	106	9	8	115	0.51
5	234	27	10	267	1.18
6	17	47	73	64	0.28
11	35	18	34	53	0.23
12	175	31	15	206	0.91

ตารางที่ 11 ข้อมูลจากภาพจำแนกประเภท โดยวิธี M.L.R.

บริเวณ สำรวจ	จำแนกเป็น		% ของตะกอน ต่อพื้นที่	จำนวนจุดภาพ ทั้งหมด	คิดเป็นขนาด พื้นที่ (กม.) ²
	น้ำ	ตะกอน			
1	194	88	31	282	1.25
2	286	47	14	333	1.47
3	55	16	22	71	0.31
4	111	12	10	123	0.54
5	233	21	8	254	1.12
6	13	46	78	59	0.26
11	29	23	44	52	0.23
12	162	39	19	201	0.89

ตารางที่ 12 ข้อมูลจากภาพจำแนกประเภท โดยวิธี M.D.S.

4.4 การเปรียบเทียบระหว่างผลที่ได้จากการจำแนกข้อมูล ซี.ซี.ที. กับรายงานการสำรวจตะกอนฯ ของ กฟผ.

มีองค์ประกอบหลายประการที่จะต้องคำนึงถึงในการเปรียบเทียบผลที่ได้จากการจำแนกฯ กับข้อมูลจากรายงานของ กฟผ. เช่น ลักษณะความแตกต่างของข้อมูลสองชนิดนี้ ซึ่งการจำแนกของเครื่องคอมพิวเตอร์เพียงแต่มองได้ว่า บริเวณหรือจุดภาพใด ๆ มีความเข้มข้นของตะกอนแฉวนลอยมากกว่าบริเวณอื่นหรือจุดภาพอื่น ๆ เท่านั้น แต่อย่างไรก็ตาม เราอาจจะหาความสัมพันธ์ระหว่างข้อมูลสองชนิดนี้ได้ นอกจากนี้ยังมีปัญหาของการกำหนดขอบเขตของบริเวณสำรวจลงบนภาพจำแนกประเภทข้อมูล (Classification Map) ปัญหาขนาดพื้นที่สำรวจฯ ต่อจำนวน sampling point ในการเฉลี่ยปริมาณความเข้มข้นของตะกอนแฉวนลอยของจุดสำรวจนั้น ๆ เป็นต้น

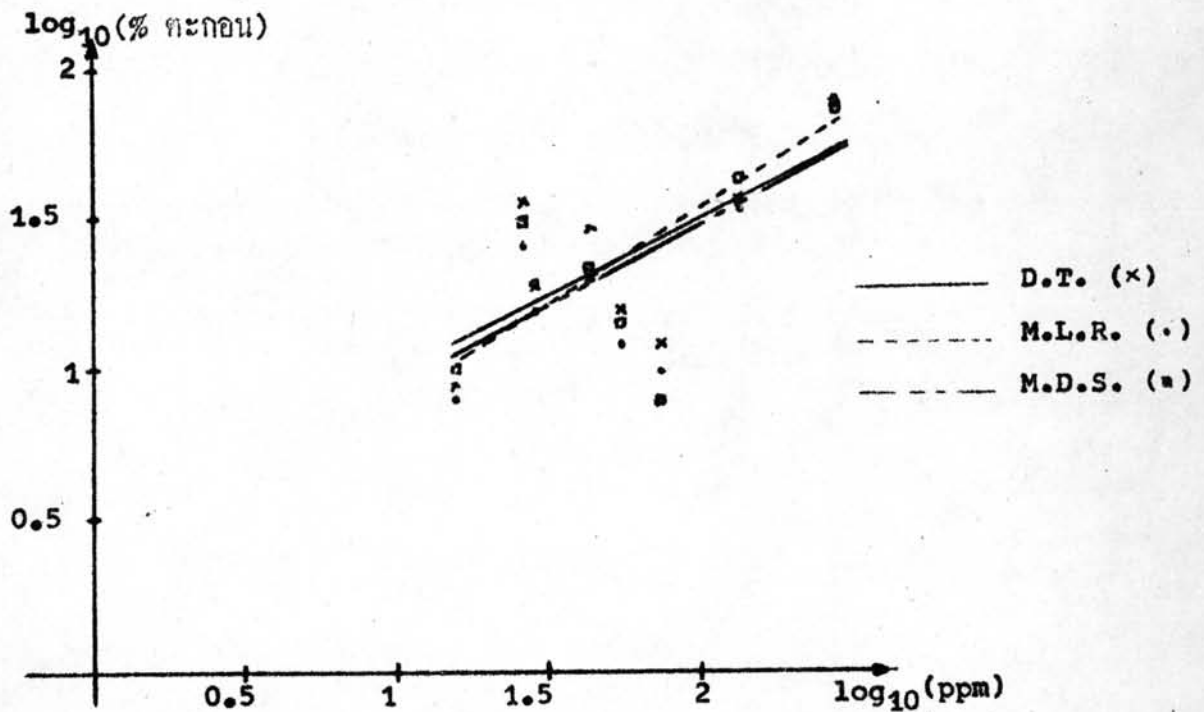
อย่างไรก็ตาม ผลที่ได้จากการจำแนกประเภทข้อมูลคือเปอร์เซ็นต์ของจำนวน
จุดภาพที่ถูกจำแนกเป็นตะกอนแขวนลอย เราจะทำการเปรียบเทียบปริมาณความเข้มข้น
เฉลี่ย หน่วยเป็นส่วนต่อล้านส่วน (ppm) ของแต่ละจุดสำรวจ ดังตารางที่ 13 ข้างล่างนี้

บริเวณ สำรวจที่	เปอร์เซ็นต์ตะกอน			ความเข้มข้นเฉลี่ย ppm	log ₁₀ (%ตะกอน)			log ₁₀ (ppm)
	D.T.	M.L.R.	M.D.S.		D.T.	M.L.R.	M.D.S.	
1	36	26	31	27	1.56	1.41	1.49	1.43
2	16	12	14	55	1.20	1.08	1.15	1.74
3	30	20	22	43	1.48	1.30	1.34	1.63
4	9	8	10	15	0.95	0.90	1.00	1.18
5	12	10	8	74	1.08	1.00	0.90	1.87
6	76	73	78	284	1.88	1.86	1.89	2.45
11	36	34	44	134	1.56	1.53	1.64	2.13
12	18	15	19	28	1.26	1.18	1.28	1.45

ตารางที่ 13 เปรียบเทียบผลที่ได้จากการจำแนกด้วยคอมพิวเตอร์กับรายงานการ
สำรวจฯ ของ กฟผ.

ตารางที่ 13 ได้ทำการคำนวณค่า logarithm ของค่าที่เป็นเปอร์เซ็นต์และ
ค่าที่เป็น ppm ไว้ด้วย ซึ่งจะเห็นได้ว่า ค่าของตัวเลขจากข้อมูลสองชุดนี้อยู่ใน order of
magnitude เดียวกัน ทำให้สะดวกต่อการนำมาเปรียบเทียบมากกว่า

ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) ระหว่างค่า
log₁₀ (% ตะกอน) และ log₁₀ (ppm) จะให้ค่าดังนี้
โดยวิธี D.T. ให้ค่า r = 0.68
M.L.R. ให้ค่า r = 0.75
M.D.S. ให้ค่า r = 0.64



รูปที่ 36 กราฟแสดงความสัมพันธ์ระหว่างค่า logarithm ของผลการจำแนก
ควยคอมพิวเตอรืกับข้อมูลจากการสำรวจของ กฟผ.

อย่างไรก็ตาม ข้อสังเกตประการหนึ่งที่ได้พบจากการสำรวจของ กฟผ. คือ จำนวนจุดที่ทำกรเก็บตัวอย่างน้ำเพื่อเฉลี่ยความเข้มข้นของปริมาณตะกอนแขวนลอย ต่อขนาดพื้นที่ของบริเวณสำรวจนั้น มีความไม่สอดคล้องอยู่บ้าง ตัวอย่างเช่น บริเวณสำรวจที่ 1 ขนาดพื้นที่ 1 ตร.กม. และ ทำกรเก็บตัวอย่างน้ำมาวิเคราะห์เพียง 10 จุด ในขณะที่บริเวณสำรวจที่ 6 ซึ่งมีขนาดพื้นที่เพียง 0.3 ตร.กม. ก็ทำการเก็บตัวอย่างมา 10 จุดเช่นเดียวกัน โดยเฉลี่ยแล้ว sampling point ของทุก ๆ บริเวณสำรวจ มีจำนวนไม่ต่างกันมากนักในขณะที่ขนาดพื้นที่ของบริเวณสำรวจต่างกันหลาย ๆ เท่า ถ้าหากเราทำการเปรียบเทียบเฉพาะบริเวณสำรวจที่มีขนาดพื้นที่ไม่เกิน 1 ตร.กม. ผลของ correlation coefficient จะเป็น 0.96, 0.97 และ 0.99 ตามลำดับ ซึ่งเป็นการยืนยันความสัมพันธ์ระหว่างข้อมูลสองชุดนี้ยิ่งขึ้น

จากค่า correlation กว่า 65% นี้ แสดงว่า ความสัมพันธ์ระหว่างผลที่ได้จากการจำแนกด้วยคอมพิวเตอร์กับรายงานของ กฟผ. อยู่ในเกณฑ์เชื่อถือได้ ดังนั้นผลการจำแนกตะกอนแขวนลอยจากข้อมูลดาวเทียมมา โดยใช้เครื่องคอมพิวเตอร์ในครั้งนี้ได้ผลดีพอสมควร

จากค่า \log (% ตะกอน) และ \log (ppm) ของแต่ละบริเวณสำรวจในตารางที่ 13 เมื่อนำไปพลอตจุดและคำนวณหา regression formula ด้วยวิธี least mean square จะได้สมการของเส้นตรง ดังแสดงไว้ข้างล่างนี้

$$y_1 = 0.495 + 0.505 x$$

$$y_2 = 0.294 + 0.570 x$$

$$y_3 = 0.434 + 0.520 x$$

โดยที่ y_1 y_2 y_3 คือค่า logarithm ของเปอร์เซ็นต์จำนวนจุลภาพที่ถูกจำแนกเป็นตะกอนแขวนลอยโดยวิธี D.T., M.L.R. และ M.D.S. ตามลำดับ
 x คือค่า logarithm ของปริมาณความเข้มข้นเฉลี่ยหน่วยเป็นส่วนต่อล้านส่วน

รูปที่ 36 เป็นกราฟแสดงความสัมพันธ์ระหว่างข้อมูล 2 ชุดนี้

นอกจากปัญหาบางประการที่ได้เอ่ยถึงในตอนต้นของหัวข้อนี้แล้ว ปัญหาสำคัญอีกประการหนึ่งที่ต้องคำนึงถึงคือ ช่วงเวลาการสำรวจกับเวลาที่ดาวเทียม ทำการบันทึกข้อมูล ซึ่งในที่นี้ได้พยายามเลือกข้อมูล ซี.ซี.ที. ที่อยู่ระหว่างกลางของช่วงเวลาสำรวจพอดี กล่าวคือ การสำรวจฯ ของ กฟผ. กระทำอยู่ในช่วงเดือนธันวาคม 2519 ถึงเดือนพฤษภาคม 2521 และข้อมูล ซี.ซี.ที. ถูกบันทึกเมื่อต้นเดือนกุมภาพันธ์ 2521

4.5 สถิติการใช้เวลาคอมพิวเตอร์ (Machine time)

สถิติเวลาที่จะได้กล่าวถึงต่อไปนี้ เป็นสถิติเวลาที่ใช้ในการวิเคราะห์งานอย่างจริงจัง โดยไม่ได้รวมถึงเวลาที่เริ่มเข้าไปทำความเข้าใจระบบโปรแกรมและทำความเข้าใจคุ้นเคยกับระบบเครื่องคอมพิวเตอร์ที่ใช้ในการวิจัยครั้งนี้ ซึ่งใช้เวลาประมาณ 20 ชม.

ตารางที่ 14 เป็นตารางสรุปการใช้เวลาคอมพิวเตอร์ (elapsed time) แต่ละขั้นตอนในการวิเคราะห์โดยประมาณ

ขั้นตอน	เวลา		หมายเหตุ
	ชม.	นาที	
1. จัดเตรียมข้อมูล	2	—	ความเร็วในการ Reformat ประมาณ 17 บรรทัด/นาที ⁽¹³⁾ โดยแต่ละบรรทัดประกอบด้วยข้อมูล 3320 ชุด
2. Pre - analysis	4	30	รวมขั้นตอนในการ DTAP, HSTGRM, DNSTRS และ PMAP
3. Analysis	9	40	รวมถึงการสร้างแฟ้มข้อมูลจาก ratioing และการคูณซึ่งใช้เวลาประมาณ 30 บรรทัด/นาที และการทำ HSTGRM, DNSTRS PMAP ฯลฯ
4. Results		36	Classification โดย D.T. และ M.D.S. ความเร็วเท่ากับการทำ PMAP คือประมาณ 200 บรรทัด/นาที ส่วนวิธี M.L.R. 2 ประเภทข้อมูล 2 แมนด ความเร็วประมาณ 10 บรรทัด/นาที

ตารางที่ 14 สรุปการใช้เวลาคอมพิวเตอร์ในการวิเคราะห์ทั้งหมด

รวมเวลาทั้งหมดประมาณ 17 ชั่วโมง อย่างไรก็ตาม ปรากฏให้เห็นได้ว่า เวลาส่วนใหญ่
หมดไปกับขั้นตอนในการวิเคราะห์โดยตรง ซึ่งเมื่อสามารถสรุปผลการวิเคราะห์ได้แล้ว
ขั้นตอนในการจำแนกประเภทข้อมูลก็จะทำได้อย่างรวดเร็วโดยใช้เวลาคอมพิวเตอร์
น้อยมาก